IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Gregory T. Bleck, et al.	Conf. No. 9065
Serial No.:	10/759,315	Group No.: 1633
Filed:	January 16, 2004	Examiner: Popa
Entitled:	**PRODUCTION OF HOST CELLS CONTAINING MULTIPLE INTEGRATING VECTORS BY SERIAL TRANSDUCTION**	

APPELLANTS' BRIEF

EFS WEB FILED
Appeal Brief - Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Examiner Popa:

This Brief is in furtherance of the Notice of Appeal filed June 16, 2011.

Appellants hereby authorize the Commissioner of the Patent and Trademark Office to charge the fees required under § 41.20(b)(2), any required fee for any Petition for Extension of Time, and any other fee for filing this Brief to Attorney Deposit Account No. 50-4302. Please reference Attorney Docket No.: GALA-08484 when charging the Attorney Deposit Account.

This Brief contains these items under the following headings and in the order set forth below [37 CFR § 1.192(c)]:

I. REAL PARTY IN INTEREST

The real party in interest is the Assignee, Catalent Pharma Solutions, LLC.

II. RELATED APPEALS AND INTERFERENCES

There are no appeals or interferences related to the pending appeal.

III. STATUS OF CLAIMS

Claims 1-42 were filed in the original application. Claims 11 and 19 were cancelled in Applicant's Response dated August 22, 2005. Claims 29 and 42 were cancelled in Applicant's Response dated February 21, 2006. Claim 13 was cancelled in Applicant's Response dated August 7, 2006. Claim 27 was cancelled in Applicant's Response dated September 4, 2007.

Claims 1-10, 12, 14-18, 20-26, 28 and 30-41 are pending and are being appealed.

Appellants appeal the Final Office Action of December 21, 2010.

The Claims, as they now stand, are set forth in Section VIII. CLAIMS APPENDIX.

IV. STATUS OF AMENDMENTS

All previous amendments have been entered.

V. SUMMARY OF CLAIMED SUBJECT MATTER

Independent Claim 1 is drawn to a method for producing a protein of interest comprising:

a) providing:

i) at least one host cell comprising a genome, wherein said host cell has an immortal phenotype, and

ii) a plurality of retroviral vectors comprising 5' and 3' long terminal repeats, wherein said retroviral vectors encode a gene of interest operably linked to an exogenous promoter internal to said 5' and 3' long terminal repeats; and

b) contacting said at least one host cell with said plurality of retroviral vectors under conditions such that said host cells are transduced to produce transduced host cells, wherein said conditions comprise contacting said host cell at a multiplicity of infection of from about 10 to 1000;

c) repeating steps a) and b) a plurality of times

d) clonally selecting a host cell expressing said gene of interest, wherein the genome of said host cell comprises from 20 to 100 integrated retroviral vectors; and

e) purifying a protein of interest encoded by said gene of interest.

This method is described in the specification generally at pages 32-48. Specifically, host cells that can be used in the methods are described at pages 33-34. Retroviral vectors that can be used with the claimed invention are described at pages 35-41. Transduction at high multiplicities of infection (MOI) is described at pages 45-48. Serial transduction procedures are described at p. 47, lines 14-29.

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

There is one grounds of rejection to be reviewed on appeal:

1. Whether Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. (Rejection 1).

2. Whether Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al., in further view of Schroder et al. (Rejection 2).

3. Whether Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. in further view of Primus et al. and Kolb et al. (Rejection 3).

4. Whether Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. in further view of Naldini et al. (Rejection 4).

VII. ARGUMENT

A. The claims are not obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. (Rejection 1).

Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are rejected under 35 U.S.C. §103, as allegedly being obvious over Mathor et al. in view of each of Burns et al., Felts et al.; Schott et al.; and Persons et al.

The Office acknowledges that: "Mathor et al. and Burns et al. do not specifically teach serial transduction to a obtain cell comprising in its genome from 20 to about 100 integrated

vectors (claims 1- 10 and 41). Final Office Action dated December 21, 2010, p. 7. The Office then alleges the following:

> However, Mathor et al. do teach that protein expression is directly proportional to the integration events (i.e., copy number) (p. 10376, column 1). Additionally, the prior art as a whole teaches that there is a positive correlation between the MOI and integration events. For example, Felts et al. teach that the advantage of retroviral vectors is that the copy number of integrated provirus can easily be controlled by varying the multiplicity of infection (MOI) (p. 74). Schott et al. teach serially transducing cells with a retroviral vector carrying an internal promoter driving the expression of a gene of interest, wherein higher MOI result in higher integration events and wherein the expression and stability of the gene of interest directly correlates with the number of integrated retroviral vectors (Abstract, p. 292, column 2, p. 294, column 2, second paragraph, p. 295, column 1, p. 302, column 2, first full paragraph, p. 303, column 2 and Fig. 9, p. 308, column 1). Persons et al. teach that repeatedly transducing cells with retroviral vectors at a MOI of 1,000 results in cells comprising 20 copies of integrated retroviral vector (Abstract, paragraph bridging p. 168 and 169, p. 171, column 2, last paragraph, p. 172, column 2, last paragraph, p. 173, column 2, p. 174, column 2, p. 177, column 2, p. 179, column 1, first full paragraph). Based on these teachings, one of skill in the art would have known that serially transducing cells with high MOI would result in increased proviral integration events. It would have been obvious to one of skill in the art, at the time the invention was made, to modify the method of Mathor et al. and Burns et al. by serially transducing their cells with high MOIs (such as MOIs of 1,000) to achieve the claimed ranges of integration events, with a reasonable expectation of success. The motivation to do so is provided by Mathor et al., who teach the possibility of specifying the level of transgene expression by controlling the integration events (Abstract, p. 10376, column 1). One of ordinary skill in the art would have been expected to have a reasonable expectation of success in doing so because the art teaches that the level of retroviral vector integration events can be easily controlled by manipulating the MOI. With respect to the limitation of an internal promoter (claim 1), using such was routine in the prior art, as taught by Schott et al. (p. 292, column 2). It would have been obvious to one of skill in the art, at the time the invention was made, to modify the method of Mathor et al. and Burns et al. by further including an internal promoter in their vector to achieve the predictable result of expressing h1L-6 in their cells. With respect to the limitations of one cell secreting more than 10 or 50 pg protein per day (claims 33 and 34), one of skill in the art would have had known to obtain the desired amounts of synthesized proteins by controlling the number of integration events.

Id. at pp. 7-9.

Applicants respectfully submit that the rejection of the claims as obvious over the cited references is inappropriate for the following reasons: 1) the Office has made scientifically unsupportable assumptions regarding the data and teachings of the primary reference, Mathor et al.; 2) the Office has failed to consider the prior art as a whole where the prior art teaches against

making cell lines with increased numbers of retroviral integrations; and 3) the Office has failed to consider Applicant's rebuttal evidence.

1. The Office has not provided a rational underpinning for the modification of Mathor et al.

A rejection for obviousness must include articulated reasoning with some rational underpinning to support the legal conclusion. *In re Kahn*, 441 F.3d 977, 988 (Fed. Cir. 2006)(Emphasis added). Applicants submit that the Office has made scientifically unsupportable assumptions regarding the data and teaching of Mathor et al. These unsupportable assumptions are at the heart of the reasons the Office provides for modifying the teaching of Mathor et al. to provide a cell line with more than integrated retroviral vectors for expression of a protein of interest. Thus, the Office failed to establish a prima facie case of obviousness because the alleged facts upon which it relies are not based on a rational underpinning. Below, Applicants explain why the Examiner's assertions concerning Mathor et al. (and the motivation to modify Mathor et al.) are not supported by articulated reasoning with rational underpinning.

At page 7 of the Final Office Action dated December 21, 2010, the Examiner states that Mathor et al. teach that protein expression is directly proportional to integration events (i.e., copy number)(p. 10376, column 1). The Examiner goes on to state that:

> "It would have been obvious to one of skill in the art, at the time the invention was made, to modify the method of Mathor et al. and Burns et al. by serially transducing their cells with high MOIs (such as MOIs of 1,000) to achieve the claimed ranges of integration events, with a reasonable expectation of success. The motivation to do so is provided by Mathor et al., who teach the possibility of specifying the level of transgene expression by controlling the integration events (Abstract, p. 10376, column 1).

Office Action, p. 8. The Examiner cites the abstract and p. 10376 of Mathor. Mathor et al. state in this section that "The rate of secretion of the exogenous protein by cultures generated by single clones was proportional to the number of integrations per progenitor cell." In sum, the Examiner is arguing that based on the results of Mathor et al., one of skill in the art would be motivated to make cells with greater than 20 integrated retroviral vectors because the rate of protein production is proportional to the number of integrations.

However, the Examiner has made assumptions about the Mathor et al. data and extrapolated the data in ways that are not scientifically supportable or appropriate. Applicants have addressed this in detail in the Fourth Declaration of Dr. Gregory Bleck (Fourth Bleck Decl.; provided in the Evidence Appendix) and the Fifth Declaration of Dr. Gregory Bleck (Fifth Bleck

Decl.; provided in the Evidence Appendix). Careful consideration of Mathor et al. and the other cited references provides the following facts:

- Mathor et al. does not contain any data or any statement that the level of transgene expression can be controlled by controlling integration events within the range of 20 to 100 integrations. See Mathor et al. in its entirety, Fifth Bleck Decl. ¶5.
- Mathor et al. presents the data on proviral integration and transgene expression on p. 10373 and in Table 1. Fifth Bleck Decl. ¶5.
- This data shows increasing transgene expression as the proviral integrations increase from 1 to 8. When the number of proviral integrations increases to 15, the transgene expression actually decreases to a level lower than was observed with 8 integrations. Id.
- Mathor et al. teaches that transgene expression correlates with number of integrations over the range of 1 to 8 integrations. Id.
- Transgene expression decreased when a cell line with 15 integrations was analyzed. Id.
- Mathor's statement that "The rate of secretion of the exogenous protein by cultures generated by single clones was proportional to the number of integrations per progenitor cell" is valid with respect to the range of 1 to 8 integrations and does not apply outside of that range. Id.
- The Examiner's attempt to apply the statement outside of the range is not factually supported, i.e., supported by the data. Fifth Bleck Decl. ¶5.
- The experiments in Mathor et al. were not conducted in a manner so that a statistical analysis could be conducted. See Mathor et al., particularly p. 10373, Fifth Bleck Decl. ¶7.
- The groups were not replicated and there is no way to determine experimental error. Id.
- Thus, it is not possible to construct a curve or equation from the data so that a correlation of transgene expression to a number of integrations outside of the data range (i.e., 20 to 100 integrations) can be made. Id.
- Any attempt to correlate the data from Mathor, which is based on unreplicated results with from 1 to 15 integrations, to the claimed 20 to 100 integrations is speculation without a factual basis. For example, based on the data in Mathor, it is speculation as to whether another clone with 15 integrations would have a level of transgene expression

that is higher or lower than the reported clone. The reason for this is that the data is not amenable to statistical analysis so that such a prediction can be made. Id.

- None of the other references relied on by the Examiner teach a correlation of transgene expression to integration number in the claimed range of 20 to 100 retroviral integrations. Fifth Bleck Decl. ¶6.
- Schott et al. teaches a correlation over the range of 1 to 9 integrations. See Schott p. 304, Fig. 9; Fifth Bleck Decl. ¶6.
- The increase between 4 and 5 integrations and 6 and 7 integrations is much greater than the increase between 7 and 9 integrations. Id.
- This is similar to the Mathor et al. data and indicates that transgene expression levels off as opposed to continuing to increase, although, predications outside of the data range cannot be validly made. Id.
- Furthermore, Lui et al. contains data on the correlation of expression of transgenes separated by an IRES and does not address transgene expression correlated to number of integrations. See Lui Abstract, Fig. 2, Gig. 4, Fig. 5; Fifth Bleck Decl. ¶7
- Stamps et al. examined the role of the T-antigen gene and its site of integration in human epithelial cell immortalization. See Stamps, p. 871, Col. 2, first full para., Fifth Bleck Decl. ¶7
- The cells examined had up to five integrations. Id.
- Stamps et al. does not comment on a correlation of transgene expression to number of integrations. Id.
- Persons et al., is directed to packaging cell lines for the production of infectious retroviral vectors, and thus is not relevant to using retroviral vectors to transduce cells to make a protein of interest. Persons et al. does not address protein production or the impact of including multiple copies of a retroviral vector in a cell line for protein production. See Persons et al., Abstract, Fourth Bleck Decl. ¶9.

The Examiner argues that one of skill in the art would know that individual clones *could* have higher expression than was observed for the 15 integrant clone in Mathor et al. As argued by the Examiner: "One of skill in the art would have known that it is statistically probable to obtain clones which, albeit having the same number of retroviral copies inserted into their

chromosomes, express different amounts of retroviral-encoded proteins." Final Office Action dated December 21, 2010, p. 13. Applicants agree that clones having the same number of integrations can express different levels of a protein. However, *the data in Mathor et al. clearly shows that expression decreased after 8 integrations*. In particular, the expression for a clone with 15 integrations was lower than expression from clones with 8 integrations. Thus, there is no basis in Mathor or the other cited references for concluding that protein expression by a single clone other than those described in the references would be substantially higher or lower on average. The data in the papers cited by the Office is simply not amenable to that type of prediction based on scientifically acceptable statistical methods. What if the clones measured in the cited references were outliers and exhibited the highest possible level of protein expression? Without more data, this conclusion is just as hard to support as the Office's conclusion that some undescribed and undocumented clones could theoretically have higher protein expression and thus a person of skill in the art would be motivated to modify Mathor et al. by inserting more retroviral vectors even though the data in Mathor et al. shows that expression peaked at 8 integrated copies and decreased when 15 copies were inserted.

In sum, the Office has failed to provide a rational underpinning for the modification of Mathor et al. to provide cells with more than 20 integrated retroviral vectors.

2. The Office has failed to consider the prior art as a whole where the prior art teaches against making cell lines with increased numbers of retroviral integrations

Applicants respectfully submit that the Office has failed to consider the prior art as a whole where the prior art teaches against making cell lines with increased numbers of retroviral integrations (i.e., more than 20 retroviral integrations). The totality of the prior art must be considered, and proceeding contrary to the accepted wisdom in the art is evidence of non-obviousness. MPEP §2145. An inference that a claimed combination would not have been obvious is especially strong where the prior art's teachings undermine the very reason being proffered as to why a person of ordinary skill would have combined the known elements. *DePuy Spine, Inc. v. Medtronic Sofamor Danek, Inc.*, 567 F.3d 1314 (Fed. Cir. 2009). This is especially applicable to the case at hand, where the prior art teaches that increasing retroviral integrations can cause inactivation of the retroviral transgene and thus decreased expression of the protein of interest. The Fourth and Fifth Bleck Decl's. identify and summarize the state of the art. These

references establish that those of skill in the art would not have been motivated to make cell lines with more than 20 integrated copies of a retroviral vector because of a concern that the vectors would inactivated by methylation. The following facts are established by these references:

- Bestor and Tycko 1996 (attached at Tab 1 to the Fifth Bleck Decl.), identify two roles of genomic methylation patterns. The first is a role of programmed demethylation and methylation during development. Fifth Bleck Decl., ¶ 8; Bestor p. 363, col. 1. The second role is that cytosine methylation is part of a genome defense system which inactivates parasitic sequences such as transposable elements and proviral DNA (i.e., integrated retroviruses). Fifth Bleck Decl., ¶ 8; Bestor p. 363, col. 1.

  This second role of methylation is directly relevant to the present invention which utilizes high levels of integrated retroviral vectors. Fifth Bleck Decl., ¶ 8. Bestor and Tycko explain this relevance:

  > The host defense hypothesis requires that the silencing apparatus recognize and inactivate parasitic sequence elements. Nearly all transposition and viral integration intermediates share certain structural features, and some satellite DNA is thought to undergo amplification by extrachromosomal rolling circle replication followed by insertion of the array into the genome. Recognition and de novo methylation of CpG sites in and around features characteristic of integration reactions would insure the inactivation of the invasive element immediately upon its integration. DNA methyltransferase may have an intrinsic ability to recognize integration intermediates that are characteristic of the above integration events. The de novo sequence specificity of mammalian DNA methyltransferase is strongly dependent on alternative secondary structures in DNA; four-way junctions in cruciform structures formed by inverted-repeats in supercoiled-plasmids are especially favored targets, as are secondary structures in artificial oligonucleotide substrates. This biochemical property suggests that invasive sequences might be targeted for de novo methylation because of their presentation of alternative secondary structures during integration (Fig la).

  (Fifth Bleck Decl., ¶ 8; Bestor p. 364, col. 2, Citations omitted).

  Thus, it was a concern that due to the nature of retroviral integration, the retroviral vectors would be targeted for inactivation by methylation. Increasing copy number enhances this problem. "A common characteristic of invasive sequences is their presence

in multiple copies, and it has recently become known that repeated sequences can interact so as to trigger their mutual silencing." Fifth Bleck Decl., ¶ 8; Bestor p. 364, col. 2. Bestor and Tycko further address retroviral vectors:

> Retroviral vectors that transducer reporter genes or therapeutic agents have been observed to undergo methylation silencing after variable periods of expression in animals. Susceptability to de novo methylation and silencing has limited the usefulness of retroviral vectors in the construction of transgenic mouse lines, and it is probable that silencing phenomena may emerge as barriers to long term somatic gene therapy in humans. Successful gene transfer may require development of delivery vectors that evade the silencing response.

Bestor, p. 365, col. 2.

Bestor and Tycko 1996 demonstrates why Gunzburg et al. is not relevant to the invention. Gunzburg et al. does not address the host defense mechanism at all or that fact that vectors had been shown to be actively silenced by methylation. Fifth Bleck Decl., ¶ 8.

- Garrick et al. 1998 (attached at Tab 2 to the Fifth Bleck Decl.) provide evidence on repeat-induced gene silencing in mammals. Fifth Bleck Decl., ¶ 8. Garrick used a lox/cre system to analyze the effect of copy number on transgene expression. Fifth Bleck Decl., ¶ 8; Garrick found that "reduction in copy number results in a marked increase in expression of the transgene and is accompanied by decreased chromatin compaction and decreased methylation at the transgene locus." Fifth Bleck Decl., ¶ 8; Garrick p. 56, col. 1., p. 58. Again, this paper provides evidence that the state of the art was that increasing copy number leads to methylation and inactivation of transgenes. Id. The vector construct was not a retroviral vector, but this data is directly relevant to inactivation of an introduced transgene. Id. As Bestor and Tycko 1996 indicated, the host defense mechanism is triggered by multiple copies of invasive sequences. Id. The transposon system used in this paper and retroviral vectors are both invasive sequences. Id.

- Cherry et al. 2000 (attached at Tab 3 to the Fifth Bleck Decl.) is co-authored by two of the leading scientists in the field, Dr. David Baltimore and Dr. Rudy Jaenisch. Fifth Bleck Decl., ¶ 8. They also recognize the role of methylation in the inactivation of

proviral genes. Id. They state:

> DNA methylation is thought to be a general mechanism used by cells to silence foreign DNA and may be involved in the cell defense against transposable elements (39). DNA methylation has also been associated with the repression of gene expression and the silencing of viral control elements (2, 14, 38). Exogenously introduced retroviruses silenced in vitro and in vivo can be reactivated by treatments that result in genome wide demethylation. In addition, transcriptionally silent endogenous retroviral elements are reactivated upon loss of genomic methylation in Dnmtl knockout mice (38). Therefore, DNA methylation is thought to causally repress expression of retroviral promoters in a variety of cell types.

Cherry, p. 7419, col. 1-2. Both methylation-dependent and methylation-independent mechanisms exist to control retroviral gene expression. Fifth Bleck Decl., ¶ 8; Cherry, p. 7425, col. 1.

- Mehtali et al. 1990 (attached at Tab 4 to the Fifth Bleck Decl.) conducted experiments that show that methylation of an introduced transgene increases with increasing copy number and that expression of the transgene decreases with increasing copy number after initially increasing. Fifth Bleck Decl., ¶ 8; See Table 1, p. 182. The vector construct was not a retroviral vector, but this data is directly relevant to inactivation of an introduced transgene. Id.

- Niwa et al. 1983 (attached at Tab 5 to the Fifth Bleck Decl.) postulated that that there are two independent mechanisms that block expression from newly acquired retroviral vectors. Fifth Bleck Decl., ¶ 8; Niwa, See Abstract, p. 1105. The first mechanism operates in undifferentiated cells to block expression of M-MuLV and other exogenously acquired viral genes, such as SV40 and polyoma virus, and does not depend on DNA methylation. Id. The second mechanism relates only to differentiated cells and represses expression of genes in which DNA is methylated. Id. This paper further serves to demonstrate why the Examiner's reliance on Gunzburg et al. is inappropriate. Fifth

Bleck Decl., ¶ 8. Newly acquired retroviral vectors are treated by cells in a different manner from proviral sequences that have been integrated into the genome in the distant past and essentially become endogenous. Id.

- Svoboda et al. 2000 (attached at Tab 6 to the Fifth Bleck Decl.) examines the expression of retroviral vectors in foreign species. Fifth Bleck Decl., ¶ 8. The vectors are subject to cell-mediated control at the transcriptional and posttranscriptional levels. Fifth Bleck Decl., ¶ 8; Svoboda, Abstract, p. 181. Of main importance is cell transcriptional regulation, which can lead to proviral silencing. Fifth Bleck Decl., ¶ 8; Svoboda p. 181, col. 2. The authors note that all of the data so far point to the important role of methylation in provirus silencing in general and that strategies for preventing methylation should contribute to more efficient gene transfer in the future. Fifth Bleck Decl., ¶ 8; Svoboda p. 186, col. 2. Again, the state of the art was that newly acquired retroviral vectors are subject to silencing by methylation. Fifth Bleck Decl., ¶ 8. This is in direct contrast to the Examiner's conclusions based on Gunzburg et al. Id.

- Ellis and Pannell 2001 (attached at Tab 7 to the Fifth Bleck Decl.) also examine retrovirus silencing. Fifth Bleck Decl., ¶ 8. They state that inclusion of appropriate regulatory elements may not be sufficient because the vectors are frequently silenced and that a better understanding of the mechanism of vector silencing is needed. Fifth Bleck Decl., ¶ 8; Ellis p. 17, col. 1-2.

- Challita and Kohn 1994 (attached at Tab 8 to the Fifth Bleck Decl.) provide data that shows that lack of expression following retroviral transduction is due to methylation. Fifth Bleck Decl., ¶ 8. As stated by the authors: "Methylation of cytosine residues has been shown to be associated with suppression of gene expression and, in certain circumstances, with the silencing of viral control elements (6). The MoMuLV-LTR is completely inactive in embryonic stem and embryonic carcinoma cell lines, and the inactivity is accompanied by de novo methylation of the proviral sequences (7, 8). Moreover, methylation has been detected in association with the MoMuLV-LTR transcriptional inactivity in fibroblasts in vitro (9) and in vivo (2)." Fifth Bleck Decl., ¶

8; Challita, p. 2567. As shown by Ellis and Pannell (Tab 6 of Fifth Bleck Decl.), these problems still had not been solved by 2001, even when regulatory elements other than the retroviral LTR are used. Fifth Bleck Decl., ¶ 8.

These references establish that at the time of the claimed invention the state of the art was that: 1) cells have a host defense mechanism that inactivates newly introduced, invading sequences such as retroviral vectors; 2) the host defense mechanism operates by methylation of the invading sequences, which causes transcriptional inactivation of the sequences; 3) transcriptional inactivation by methylation leads to reduced expression from retroviral vectors; 4) the inactivation may be triggered by structures formed during integration of the retroviral vectors; and 5) the presence of multiple repeats of an invading sequence such as a retroviral vector triggers methylation and inactivation. Fifth Bleck Decl., ¶ 9. The prior art, when considered as a whole, teaches away from the present invention. The prior art teaches that increasing retroviral integrations can cause inactivation of the retroviral transgene and thus decreased expression of the protein of interest. This supports the conclusion that the claimed invention is not obvious.

The Office cited Gunzburg et al. in an attempt to rebut Applicant's evidence. Office Action dated May 13, 2010, p. 15. Dr. Bleck addressed Gunzburg in his Fifth declaration. Gunzburg et al., which was published in 1984, addresses methylation of "multiple endogenous mouse mammary tumour virus (MMTV) proviral genes" that "are present at different locations in mouse inbred strains." See Abstract, Fifth Bleck Decl., ¶ 8. Gunzburg et al. finds that the methylation patterns are location and tissue specific and that the patterns are stably inherited and appear to be conferred upon the viral DNA by the flanking mouse genomic DNA. Id.. The authors state that "upon integration the provirus assumes the methylation pattern of the DNA into which is integrates." p. 1129, col. 1., p. 133, col. 2, Fifth Bleck Decl., ¶ 8. Importantly, Gunzburg et al. does not contain data or comments that address any correlation of methylation to expression of genes. Fifth Bleck Decl., ¶ 8. Just as important, Gunzburg et al. addresses endogenous proviral sequences and not the introduction of exogenous vectors containing transgenes. Fifth Bleck Decl., ¶ 8 These proviral sequences are endogenous to the genome and have been acquired at some point in the distant past. Fifth Bleck Decl., ¶ 8 Gunzburg et al. has

very little relevance to the present invention or to the evidence previously submitted by Dr. Bleck.

In response, the Office argued as follows:

> The applicant argues that, since Gunzburg et al. relate to endogenous and not exogenous proviral sequences and since they do not show data nor do they provide comments addressing any correlation of methylation and gene expression, their teachings have very little relevance to the instant invention. The applicant also argues that neither Stamps et al. nor Liu et al. teach a correlation between copy number and transgene expression. These arguments are not material to the instant rejection because the references were no used to reject the claims and because the references were not cited for providing data correlation copy number with transgene expression. The references were only cited as evidence that the existence of a correlation between the insertion site and transgene expression level was common knowledge in the prior art. Furthermore, Gunzburg et al. do teach that there is a correlation between the insertion site and expression of retroviral genes for both endogenous and exogenous proviral sequences, including MoMLV (p. 1129, bridging columns 1 and 2; paragraph bridging p. 1133 and 1134; p. 1134, column 1). Thus, the teachings of Gunzburg et al. are relevant to the instant invention.

These arguments do not address the fact that Gunzburg, published in 1984 (more than 16 years before the priority date of the instant invention) provides an analysis of methylation of endogenous proviral genes that have been acquired at some point in the distant past. Gunzburg says nothing about gene silencing by methylation and thus does nothing to rebut the facts established by the references cited in the Fifth Bleck Decl.

In conclusion, the evidence of record supports the fact that the prior art taught away from introducing more than 20 retroviral vectors encoding a protein of interest into a cell to produce the protein of interest because the prior art predicted that such vectors would be inactivated by methylation, thus decreasing production of the protein of interest. The combination and/or modification of Mathor et al. in view of the remaining cited references is therefore not obvious as argued by the Examiner.

The Examiner has also failed to consider Applicants evidence concerning references in the prior art that teach away from the invention as claimed. Applicants have cited multiple references that establish that at the time of the claimed invention the state of the art was that: 1) cells have a host defense mechanism that inactivates newly introduced, invading sequences such as retroviral vectors; 2) the host defense mechanism operates by methylation of the invading sequences, which causes transcriptional inactivation of the sequences; 3) transcriptional inactivation by methylation leads to reduced expression from retroviral vectors; 4) the inactivation may be triggered by structures formed during integration of the retroviral vectors; and 5) the presence of multiple repeats of an invading sequence such as a retroviral vector triggers methylation and inactivation. Fifth Bleck Decl., ¶ 8-9; See Response dated October 7, 2010 p. 12-18 for a complete discussion.[1]

These facts establish the state of the art. The Examiner has failed to consider or give proper weight to this evidence.

3. The Office has failed to consider Applicant's rebuttal evidence.

As discussed above, Applicants have submitted a substantial amount of evidence in rebuttal of the Examiner's alleged prima facie of obviousness. When a *prima facie* case is made (for the record, Applicants do not believe a *prima facie* case has been made here), the burden shifts to the applicant to come forward with evidence and/or argument supporting patentability. *In re Sullivan*, 498 F. 3d 1345, 1351, 81 USPQ2d 1034 (Fed. Cir. 2007) (citing *In re Glaug*, 283 F.3d 1335, 1338 (Fed.Cir.2002)). As held in *Sullivan*:

> Rebuttal evidence is "merely a showing of facts supporting the opposite conclusion." *In re Piasecki*, 745 F.2d 1468, 1472 (Fed.Cir.1984). Evidence rebutting a *prima face* case of obviousness can include: "evidence of unexpected results," *Pfizer, Inc. v. Apotex, Inc.*, 480 F.3d 1348, 1369 (Fed.Cir.2007), evidence "that the prior art teaches away from the

[1] See also Bestor and Tycko 1996 (Tab 1 to the Fifth Bleck Decl.), Garrick et al. 1998 (Tab 2 to the Fifth Bleck Decl.); Cherry et al. 2000 (Tab 3 to the Fifth Bleck Decl.); Mehtali et al. 1990 (Tab 4 to the Fifth Bleck Decl.); Niwa et al. 1983 (Tab 5 to the Fifth Bleck Decl.); Svoboda et al. 2000 (Tab 6 to the Fifth Bleck Decl.); Ellis and Pannell 2001 (Tab 7 to the Fifth Bleck Decl.); Challita and Kohn 1994 (attached at Tab 8 to the Fifth Bleck Decl.))

claimed invention in any material respect," *In re Peterson*, 315 F.3d 1325, 1331 (Fed.Cir.2003), and evidence of secondary considerations, such as commercial success and long-felt but unresolved needs, *WMS Gaming, Inc. v. Int'l Game Tech.*, 184 F.3d 1339, 1359 (Fed.Cir.1999).

Id. Importantly, when a patent applicant puts forth rebuttal evidence, the Office must consider that evidence. *Id., see also In re Soni*, 54 F.3d 746, 750 (Fed.Cir.1995) (stating that "all evidence of nonobviousness must be considered when assessing patentability"); *In re Sernaker*, 702 F.2d 989, 996 (Fed.Cir.1983) ("If, however, a patent applicant presents evidence relating to these secondary considerations, the board must always consider such evidence in connection with the determination of obviousness.").

The determination whether an invention would have been obvious under 35 U.S.C. § 103 is a legal conclusion based on underlying findings of fact. *Id.* at 1350 (citing *In re Kotzab*, 217 F.3d 1365, 1369 (Fed.Cir.2000)). With respect to the instant invention, the Office has failed to establish a *prima facie* case evidence because the facts the Office has relied on are scientifically incorrect or not supported by the references cited by the Office and, in any event, the Office has failed to considered properly established facts submitted in rebuttal. Instead of considering the evidence in the Fourth and Fifth Bleck Decl's., the Office has merely maintained the same rejections over the same prior art. In doing so, the Office has relied on references such Gunzburg et al. in an attempt to address Applicant's evidence. However, as explained above those references and arguments are not relevant to the evidence presented by Applicants.

The weight of the evidence presented by Applicants, and not properly considered or addressed by the Office, establishes that the claimed invention is not obvious.

B. Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are not obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al., in further view of Schroder et al. (Rejection 2).

The additional citation of Schroder does not cure the defects noted for the rejection over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. discussed in Section A above. Namely, Applicants submit that 1) the Office has made scientifically unsupportable assumptions regarding the data and teachings of the primary reference, Mathor et al.; 2) the Office has failed to consider the prior art as a whole where the prior art teaches against making cell lines with increased numbers of retroviral integrations; and 3) the Office has failed to consider Applicant's rebuttal evidence. These same arguments apply to this rejection.

C. Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are not obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. in further view of Primus et al. and Kolb et al. (Rejection 3).

The additional citations of Primus and Kolb does not cure the defects noted for the rejection over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. discussed in Section A above. Namely, Applicants submit that 1) the Office has made scientifically unsupportable assumptions regarding the data and teachings of the primary reference, Mathor et al.; 2) the Office has failed to consider the prior art as a whole where the prior art teaches against making cell lines with increased numbers of retroviral integrations; and 3) the Office has failed to consider Applicant's rebuttal evidence. These same arguments apply to this rejection.

D. Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 are not obvious over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. in further view of Naldini et al. (Rejection 4).

The additional citation of Naldini does not cure the defects noted for the rejection over Mathor et al. in view of each of Burns et al., Felts et al., Schott et al., and Persons et al. discussed in Section A above. Namely, Applicants submit that 1) the Office has made scientifically unsupportable assumptions regarding the data and teachings of the primary reference, Mathor et al.; 2) the Office has failed to consider the prior art as a whole where the prior art teaches against making cell lines with increased numbers of retroviral integrations; and 3) the Office has failed to consider Applicant's rebuttal evidence. These same arguments apply to this rejection.

VIII. CLAIMS APPENDIX

1. (Previously presented) A method for producing a protein of interest comprising:

a) providing:

i) at least one host cell comprising a genome, wherein said host cell has an immortal phenotype, and

ii) a plurality of retroviral vectors comprising 5' and 3' long terminal repeats, wherein said retroviral vectors encode a gene of interest operably linked to an exogenous promoter internal to said 5' and 3' long terminal repeats; and

b) contacting said at least one host cell with said plurality of retroviral vectors under conditions such that said host cells are transduced to produce transduced host cells, wherein said conditions comprise contacting said host cell at a multiplicity of infection of from about 10 to 1000;

c) repeating steps a) and b) a plurality of times

d) clonally selecting a host cell expressing said gene of interest, wherein the genome of said host cell comprises from 20 to 100 integrated retroviral vectors; and

e) purifying a protein of interest encoded by said gene of interest.

2. (Original) The method of Claim 1, wherein steps a and b are repeated at least 3 times.

3. (Original) The method of Claim 1, wherein steps a and b are repeated at least 4 times.

4. (Original) The method of Claim 1, wherein steps a and b are repeated at least 5 times.

5. (Original) The method of Claim 1, wherein steps a and b are repeated at least 6 times.

6. (Original) The method of Claim 1, wherein steps a and b are repeated at least 7 times.

7. (Original) The method of Claim 1, wherein steps a and b are repeated at least 8 times.

8. (Original) The method of Claim 1, wherein steps a and b are repeated at least 10 times.

9. (Original) The method of Claim 1, wherein steps a and b are repeated at least 20 times.

10. (Original) The method of Claim 1, wherein steps a and b are repeated between about 3 and 20 times.

11. (Cancelled).

12. (Previously presented) The method of Claim 1, wherein said retroviral vectors utilized in steps a and b are produced from packaging cells transfected with an envelope plasmid and a vector plasmid.

13. (Cancelled).

14. (Original) The method of Claim 12, wherein said packaging cells express retroviral gag and pol proteins.

15. (Original) The method of Claim 14, wherein said packaging cells are 293-GP cells.

16. (Original) The method of Claim 12, wherein said envelope plasmid encodes a G protein.

17. (Original) The method of Claim 16, wherein said G protein is VSV-G protein.

18. (Original) The method of Claim 1, wherein said retroviral vector comprises MoMLV elements.

19. (Cancelled).

20. (Original) The method of Claim 1, wherein said gene of interest is operably linked to an exogenous promoter.

21. (Previously presented) The method of Claim 1, wherein gene of interest is operably linked to a secretion signal sequence.

22. (Original) The method of Claim 1, wherein said retroviral vector encodes at least two genes of interest.

23. (Original) The method of Claim 22, wherein said at least two genes of interest are arranged in a polycistronic sequence.

24. (Original) The method of Claim 23, wherein said at least two genes of interest comprise immunoglobulin heavy and light chains.

25. (Original) The method of Claim 1, wherein said retroviral vector is a lentiviral vector.

26. (Previously presented) The method of Claim 1, wherein said host cell is selected from Chinese hamster ovary cells-and human 293 cells.

27. (Cancelled).

28. (Previously presented) The method of Claim 1, further comprising culturing said clonally selected host cells under conditions such that a protein of interest encoded by said gene of interest is produced.

29. (Cancelled).

30. (Original) The method of Claim 28, further comprising isolating said protein of interest.

31. (Original) The method of Claim 28, wherein said culture conditions are selected from the group consisting of roller bottle cultures, perfusion cultures, batch fed cultures, and petri dish cultures.

32. (Previously presented) The method of Claim 28, wherein said clonally selected host cells synthesize greater than about 1 picograms per cell per day of said

protein of interest.

33. (Previously presented) The method of Claim 28, wherein said clonally selected host cells synthesize greater than about 10 picograms per cell per day of said protein of interest.

34. (Previously presented) The method of Claim 28, wherein said clonally selected host cells synthesize greater than about 50 picograms per cell per day of said protein of interest.

35. (Original) The method of Claim 1, wherein said retroviral vector further encodes an amplifiable marker.

36. (Original) The method of Claim 35, wherein said amplifiable marker is selected from the group consisting of DHFR and glutamine synthetase.

37. (Original) The method of Claim 35, further comprising the step of culturing said transduced host cells under conditions that allow for amplification of the integrated retroviral vectors.

38. (Original) The method of Claim 37, wherein said conditions comprise culturing said transduced host cells in the presence of a selection agent selected from the group consisting of methotrexate, phosphinothricin and methionine sulphoxime.

39. (Previously presented) The method of Claim 24, wherein said immunoglobulins are selected from the group consisting of IgG, IgA, IgM, IgD, IgE and sIg.

40. (Original) The method of Claim 1, wherein said host cell is transduced with at least two different vectors encoding different genes of interest.

41. (Previously presented) A host cell produced by the method of Claim 1, wherein said host cell comprises from 20 to about 100 integrated retroviral vectors.

42. (Cancelled).

IX. EVIDENCE APPENDIX

This evidence appendix contains the following Declarations and associated exhibits:

1. Fourth Declaration of Dr. Gregory Bleck, entered by the Examiner in the Office Action mailed May 13, 2010.

2. Fifth Declaration of Dr. Gregory Bleck, entered by the Examiner in the Office Action mailed December 21, 2010.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Gregory T. Bleck, et al.	
Serial No.:	10/759,315	Group No.: 1633
Filed:	1/16/04	Examiner: Popa

Entitled: **PRODUCTION OF HOST CELLS CONTAINING MULTIPLE INTEGRATING VECTORS BY SERIAL TRANSDUCTION**

FOURTH DECLARATION OF DR. GREGORY BLECK

I, Dr. Gregory Bleck, state as follows:

1. My present position is Senior Director, Cell Line Engineering, Catalent Pharma Solutions-Middleton, WI.

2. I am an inventor of the above referenced patent application.

3. In the Office Action dated July 17, 2009, the Examiner addressed my previous Declarations by stating:

> Applicant argues that the Second Bleck Declaration establishes that one of skill in the art would not extrapolate the data of Mathor et al. to conclude that making cell lines with genomes comprising greater than 20 integration events was desired or feasible. The Second Bleck Declaration was previously presented and addressed. Furthermore, apart from an argument, Applicant and his Declaration did not provide any evidence indicating that obtaining host cells with genomes comprising greater than 20 integrated viral vectors was not feasible before the instant invention was made. It is noted that Table 1 (to which Applicant refers to) shows the result obtained with only one keratinocyte clone for each 8 and 15 integration events. Such results cannot be extrapolated to all clones. The Examiner provided evidence to support this statement and it is not clear why Applicant argues that such a statement is sheer speculation and an unsupported extrapolation of the data. It is well established in the art that retroviral insertion is random and that expression level is dependent on the insertion sites (see Mathor et al., of record, p. 10376, column 1; Liu et al., of record, Anal Biochem, 2000, 280: 20-28, Abstract, p. 21, column 1; Stamps et al, of record, Int J Cancer, 1994, 57: 865-874, Abstract, p. 868, column 1, p. 869, Fig. 2). It is noted that Applicant did not indicate why the teachings of Liu et al. and Stamps et al. do not support the Examiner's statement. Applicant points to Table 1 in Mathor et al. for support. In fact, Table 1 in Mathor et al. clearly demonstrates that results obtained with

one clone cannot be extrapolated to other clones. Specifically, Table 1 demonstrates that three different clones each comprising 3 copies of integrated vectors secrete different amounts of hIL-6 (180, 150, and 450 ng/10^6 cells/day, respectively); two different clones comprising 4 copies of integrated vectors also secrete different amounts of hIL-6 (522 and 449 ng/106 cells/day, respectively). Therefore, the Examiner's statement is supported by the art, including the data in Table 1 of Mathor et al. Based on these teachings, one of skill in the art would expect clones with the same number of copies to have different expression levels, and therefore, would know to select multiple clones and look for the high expressing ones. In fact, as indicated in the rejection above, the prior art does teach obtaining stable host cells comprising 20 integrated copies of retroviral vectors. The prior art also teaches that the number of integrating events can be controlled by varying the MOI and that gene expression directly correlates with the number of integrated copies (see the rejection above). Based on the teachings in the art, one of skill in the art would have known how to introduce 20 or more copies of retroviral vector into a host cell genome. Therefore, the Examiner did not ignore the actual data and the statement that it would have been obvious and within the capabilities of one of skill in the art to obtain host cells with 20 or more integrated copies is not sheer speculation or unsupported extrapolation of the data by the Examiner.

Applicant argues that, since Zielske teaches that use of internal promoters such as CMV results in a plateau of expression at 4 copies per cell, is unexpected in view of the art. This is not found persuasive. Zielske teaches that the limit in transgene expression is related to the particular vector/promoter/transgene/host cell system used in his experiments (p. 926, paragraph bridging columns 1 and 2, p. 929, column 1, third paragraph). In support of these teachings, Schott et al. disclose that protein expression from retroviral vectors comprising an internal CMV promoter does not reach a plateau when increasing the copy number above 4. Moreover, the claims are not limited to the CMV promoter; they recite any internal promoter. The prior art teaches successful expression from internal promoters in general, including CMV, and therefore the claimed invention is not unexpected.

4. The Examiner first states that "apart from an argument, Applicant and his Declaration did not provide any evidence indicating that obtaining host cells with genomes comprising greater than 20 integrated viral vectors was not feasible before the instant invention was made." This simply is not true. The Declarations provided citations to papers that demonstrate why it was completely unexpected that stable, high-producing cell lines could be developed that contained greater than 20 integrated retroviral vectors. In further support of the fact that the claimed invention was not expected and was not predictable, I am attaching 18 scientific papers to this Declaration that demonstrate why, at the time we made this invention, it was unexpected and not predictable that making cells lines with greater than 20 integrated retroviral vectors would be successful. These papers demonstrate that at the time of the invention, the usefulness of retroviral vectors for expressing proteins was in great question because expression from the

vectors was often inactivated by methylation. The following references demonstrate this:

AKGUN et al, "Determinants of Retrovirus Gene Expression in Embryonal Carcinoma Cells," Journal of Virology, Jan. 1991, Vol. 65(1) P. 382-388

BESTOR TH et al, "Creation of genomic methylation patterns," Nature Genetics, 1996, Vol. 12(4) P. 363-7

CHALLITA et al, "Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo," Proc. Natl. Acad Sci USA, March 1994, Vol. 91, P. 2567-2571

CHERRY et al, "Retroviral Expression in Embryonic Stem Cells and Hematopoietic Stem Cells," Molecular and Cellular Biology, Oct. 2000, Vol. 20(20) P. 7419-7426

ELLIS et al, "The beta-globin locus control region versus gene therapy vectors: a struggle for Expression" Clinical Genetics, Jan. 2001, Vol. 59(1) P. 17-24

GORMAN et al, "Negative Regulation of Viral Enhancers in Undifferentiated Embryonic Stem Cells," Cell, Sept. 1985, Vol. 42, P. 519-526

HOEBEN et al, "Inactivation of the Moloney Murine Leukemia Virus Long Terminal Repeat in Murine Fibroblast Cell Lines Is Associated with Methylation and Dependent on Its Chromosomal Position," Journal of Virology, Feb. 1991, Vol. 65(2), P. 904-912

JOLLY et al, "Variable Stability of a Selectable Provirus after Retroviral Vector Gene Transfer into Human Cells," Molecular and Cellular Biology, Apr. 1986, Vol. 6(4), P. 1141-1147

PALMER et al, "Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes," Proc. Natl. Acad. Sci. USA, Feb. 1991, Vol. 88, P. 1330-1334

PANNELL et al, "Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes," The EMBO Journal, 2000, Vol. 19(21), P. 5884-5894

PANNELL et al, "Silencing of gene expression: implications for design of retrovirus vectors," Reviews in medical Virology, 2001, Vol. 11, P. 205-217

NIWA et al, "Independent Mechanisms Involved in Suppression of the Moloney Leukemia Virus Genome during Differentiation of Murine Teratocarcinoma Cells," Cell, Apr. 1983, Vol. 32, P. 1105-1113

SINGAL et al, "Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells," Proc. Natl. Acad. Sci. USA, Dec. 1997, Vol. 94, P. 13724-13729

SVOBODA et al, "Retroviruses in foreign species and the problem of provirus silencing," Gene, 2000, Vol. 261, P. 181-188

TURKER et al, "Formation of methylation patterns in the mammalian genome," Mutation Research, 1997, Vol. 386(2), P. 119-130

WANG et al, "High-Resolution Analysis of Cytosine Methylation in the 5' Long Terminal Repeat of Retroviral Vectors," Human Gene Therapy, Nov. 1998, Vol. 9(16), P. 2321-2330

Furthermore, the references indicate that the problem with inactivation was associated with increased copy number as demonstrated by these references:

GARRICK et al, "Repeat-induced gene silencing in mammals," Nature Genetics, Jan. 1998, Vol. 18(1), P. 56-9

METHTALI et al, "The methylation-free status of a housekeeping transgene is lost at high copy number," Gene, 1990, Vol. 91(2), P. 179-84

These problems were observed across cell types, including stem cells, hematopoietic stem cells, fibroblasts and teratocarcinoma cells. Thus, at the time of the invention, it would have not have been expected or predictable that making cells that contain greater than 20 copies of a retroviral vector would be useful expressing proteins. At the time of the invention, based on the references cited above, it would have been expected that increasing the copy number of retroviral vectors past 20 would result in methylation, inactivation and instability.

5. Then the Examiner supports her conclusions that the claimed invention is expected by making arguments that mischaracterize the cited references and interpret and extrapolate the data in ways that are scientifically improper.

6. First, the Examiner states that:

> It is noted that Table 1 (to which Applicant refers to) shows the result obtained with only one keratinocyte clone for each 8 and 15 integration events. Such results cannot be extrapolated to all clones. The Examiner provided evidence to support this statement and it is not clear why Applicant argues that such a statement is sheer speculation and an unsupported extrapolation of the data. It is well established in the art that retroviral insertion is random and that expression level is dependent on the insertion sites (see Mathor et al., of record, p. 10376, column 1; Liu et al., of record, Anal Biochem, 2000, 280: 20-28, Abstract, p. 21, column 1; Stamps et al., of record, Int J Cancer, 1994, 57: 865-874, Abstract, p. 868, column 1, p. 869, Fig. 2). It is noted that Applicant did not

> indicate why the teachings of Liu et al. and Stamps et al. do not support the Examiner's statement. Applicant points to Table 1 in Mathor et al. for support. In fact, Table 1 in Mathor et al. clearly demonstrates that results obtained with one clone cannot be extrapolated to other clones. Specifically, Table 1 demonstrates that three different clones each comprising 3 copies of integrated vectors secrete different amounts of hll-6 (180, 150, and 450 ng/10^6 cells/day, respectively); two different clones comprising 4 copies of integrated vectors also secrete different amounts of hll-6 (522 and 449 ng/10^6 cells/day, respectively). Therefore, the Examiner's statement is supported by the art, including the data in Table 1 of Mathor et al. Based on these teachings, one of skill in the art would expect clones with the same number of copies to have different expression levels, and therefore, would know to select multiple clones and look for the high expressing ones.

The issue here is what Table 1 of Mathor shows to one of skill in the art and in particular whether the data in Mathor indicates to one of skill in the art that cells lines with greater than 20 integrated retroviral vectors would yield higher expression levels than cells lines with 8 or 15 integrated retroviral vectors. The data in Table 1 of Mathor, which is limited to a maximum of 15 integrations, cannot be extrapolated to a situation where there are 20 integrations. It is impossible to do a valid statistical analysis or curve fit based on the data in Mathor. To provide any other interpretation to the data is not scientifically correct. For example, as shown in Appendix 1, the data could indicate a plateau or upside-down U shaped curve as shown. This is why I have previously said it was not proper to extrapolate the data to cells that have 20 integrations. A person of skill in the art would not do this.

In response, the Examiner has argued that "It is well established in the art that retroviral insertion is random and that expression level is dependent on the insertion sites (see Mathor et al., of record, p. 10376, column 1; Liu et al., of record, Anal Biochem, 2000, 280: 20-28, Abstract, p. 21, column 1; Stamps et al., of record, Int. J Cancer, 1994, 57: 865-874, Abstract, p. 868, column 1, p. 869, Fig. 2). While this may be true, it is not relevant to why the data of Mathor can be extrapolated to support the Examiner's conclusion and what that data indicates to one of skill in the art. The person of skill in the art would interpret that data as shown in Appendix 1, especially in light of the knowledge of methylation and gene silencing, as indicating that expression was declining or plateauing with increasing copy number. In particular, neither Lui nor Stamps provide any information that can be used to predict whether a cell line containing 20 integrated retroviral vectors will produce more protein than a cell line with 8 or 15 integrated retroviral vectors. For this reason, citation of Liu and Stamps provides information which is not scientifically relevant and do not support a scientifically valid argument.

The Examiner further relies on data in Mathor that shows that "three different clones each

comprising 3 copies of integrated vectors secrete different amounts of hIL-6 (180, 150, and 450 ng/10^6 cells/day, respectively); two different clones comprising 4 copies of integrated vectors also secrete different amounts of hIL-6 (522 and 449 ng/10^6 cells/day, respectively). It is certainly expected that different cell lines with the same number of integrations would have different expression levels. However, this has no relevance or predictive value with respect to the issue of whether, according to the data in Table 1, a cell line with 20 integrated retroviral vectors would produce more protein than a cell line with 8 or 15 integrated vectors. Indeed, based on the data, the curve appears to be declining or at best plateauing. No other prediction can be made from the data.

7. The Examiner further states that:

> Applicant argues that, since Zielske teaches that use of internal promoters such as CMV results in a plateau of expression at 4 copies per cell, is unexpected in view of the art. This is not found persuasive. Zielske teaches that the limit in transgene expression is related to the particular vector/promoter/transgene/host cell system used in his experiments (p. 926, paragraph bridging columns 1 and 2, p. 929, column 1, third paragraph).

The Examiner's argument that Zielske teaching are limited to a particular vector/promoter/transgene system is also not relevant to the issue of whether a person of skill in the art would have predicted or expected that that cells lines with greater than 20 integrated retroviral vectors would yield higher expression levels than cells lines with 8 or 15 integrated retroviral vectors. The Examiner initially cited Zielske as support for an argument that the person of skill in the art would have been motivated to make cell lines containing multiple copies of an integrated retroviral vector. However, Zielske does not support making cell lines with 20 integrated vectors regardless of the vector/promoter/transgene system. I note that Mathor also teaches a plateau or expression maximum at 8 integrated copies with a different vector/promoter/transgene system. Thus, Mathor and Zielske have similar teaching that would discourage making cell lines that contain greater than 20 integrated retroviral vectors.

8. The Examiner next argues that:

> In support of these teachings, Schott et al. disclose that protein expression from retroviral vectors comprising an internal CMV promoter does not reach a plateau when increasing the copy number above 4. Moreover, the claims are not limited to the CMV promoter; they recite any internal promoter. The prior art teaches successful expression from internal promoters in general, including CMV, and therefore the claimed

invention is not unexpected.

Schott does not teach a plateau at 4 copies, but only includes data to 9 copies. Again Schott is not relevant to the issue of whether a person of skill in the art would have predicted or expected that that cells lines with greater than 20 integrated retroviral vectors would yield higher expression levels than cells lines with 8 or 15 integrated retroviral vectors. Based on a combination of Zielske, Mathor and the other references I provided above, one would expect expression to decline or plateau.

9. The Examiner further argues that:

> In fact, as indicated in the rejection above, the prior art does teach obtaining stable host cells comprising 20 integrated copies of retroviral vectors. The prior art also teaches that the number of integrating events can be controlled by varying the MOI and that gene expression directly correlates with the number of integrated copies (see the rejection above). Based on the teachings in the art, one of skill in the art would have known how to introduce 20 or more copies of retroviral vector into a host cell genome. Therefore, the Examiner did not ignore the actual data and the statement that it would have been obvious and within the capabilities of one of skill in the art to obtain host cells with 20 or more integrated copies is not sheer speculation or unsupported extrapolation of the data by the Examiner.

The reference cited by the Examiner to show 20 integrated copies, Persons et al., is directed to packaging cell lines for the production of infectious retroviral vectors, and thus is not relevant to using retroviral vectors to transduce cells to make a protein of interest. Persons et al. does not not address protein production or the impact of including multiple copies of a retroviral vector in a cell line for protein production.

10. The Examiner also argues that the teachings of Bestor relate to silencing in vivo, not in vitro. This statement is untrue as many of the references cited by Bestor and known and available to those of skill in the art clearly show this effect in vitro. In this regard, the Examiner is referred to the references cited above in Paragraph 4. Furthermore, the Examiner's arguments based on Persons (i.e., 20 copies of a vector in the genome with no silencing) are not relevant to the claims because Persons is directed to making retroviral packaging cells that produce infectious retroviral particles, not cell lines that are used for protein production.

11. I further declare that all statement made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of title 18 of the United States

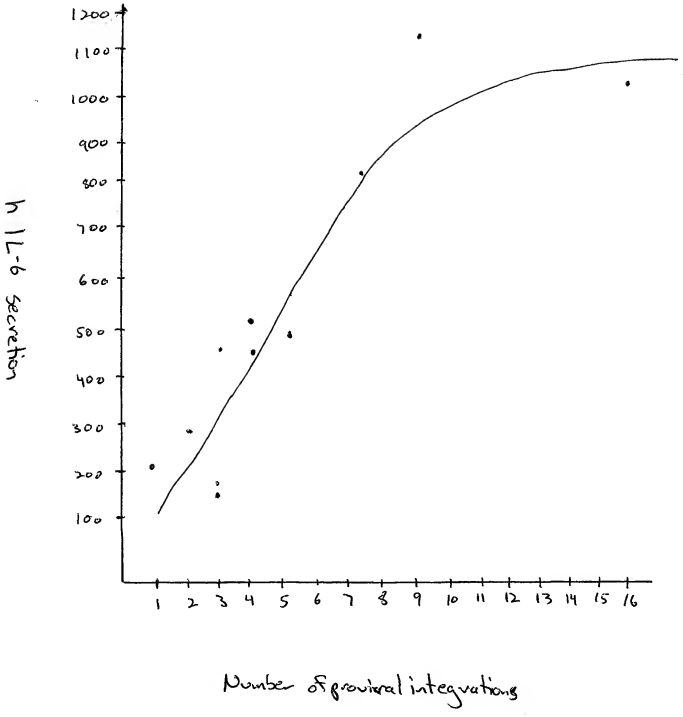
Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Dated: September 17, 2009

Dr. Gregory Bleck

APPENDIX
TABLE OF CONTENTS

TAB 1


IL-6 secretion
1200
1100
1000
900
800
700
600
500
400
300
200
100
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Number of proviral integrations

TAB 2

JOURNAL OF VIROLOGY, Jan. 1991, p. 382–388
0022-538X/91/010382-07$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 65, No. 1

Determinants of Retrovirus Gene Expression in Embryonal Carcinoma Cells

ERCAN AKGÜN, MARION ZIEGLER, AND MANUEL GREZ†*

Abteilung für Zellbiologie, Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie, Martinistrasse 52, 2000 Hamburg 20, Federal Republic of Germany

Received 5 April 1990/Accepted 12 October 1990

The expression of Moloney murine leukemia virus is restricted in embryonal carcinoma (EC) cells. To characterize specific mutations necessary for expression of retroviruses in EC cells, we analyzed the expression of retrovirus mutants and recombinants thereof in EC cell lines F9 and PCC4. DNA sequence comparison and functional studies allowed us to define three point mutations in the enhancer region of the viral mutants at positions −345, −326, and −166 and two point mutations within the 5′-untranslated region of the viral genome at positions +164 and +165 that were essential for retrovirus expression in EC cells. DNA fragments derived from either the wild type or mutant viruses were used to search for sequence-specific DNA-binding factors in nuclear extracts from undifferentiated PCC4 cells. A cellular factor was found to bind strongly to sequences within the enhancer region (−354 to −306) of wild-type viruses but only weakly to sequences derived from mutant viruses. This factor was named ECF-I (for EC cell factor I). Retroviral expression in EC cells correlates with decreased binding affinity for ECF-I.

Murine embryonal carcinoma (EC) cells, the multipotential stem cells of teratocarcinomas, are extensively used as a model system to study gene regulation in early embryonic development. EC cells share many properties with early embryonic stem cells, including restriction to retrovirus replication (10, 16, 28). Expression of Moloney murine leukemia virus (Mo-MuLV) in EC cells is barely detectable, although retroviral infection and integration occur normally in these cells (5, 18, 20, 24, 26, 29). Restriction to retroviral expression in undifferentiated EC cells is, at least in part, due to the impaired function of the Mo-MuLV enhancer in these cells (12). Lack of transcriptional activators and the presence of repressors of enhancer activity have been proposed to explain the absence of long terminal repeat (LTR)-mediated transcription in EC cells (3, 7, 13, 14). A second region essential for virus expression in EC cells is located 3′ from the LTR and includes sequences spanning the proline tRNA primer-binding site (1, 3, 13, 14). The presence of these sequences reduces the steady-state levels of RNA expressed from 5′-positioned promoters in EC cells but not in differentiated cells (1, 3, 13, 14, 33).

Retroviral mutants that are expressed in EC cells have been described (1, 4, 9, 22). One such mutant is the myeloproliferative sarcoma virus (MPSV) which was isolated from adult mice after serial transplantation of tumors induced in newborn mouse by Moloney murine sarcoma virus (Mo-MuSV; 11, 19). The enhancer region of MPSV contains several point mutations which have been shown to be responsible for expansion of the host range of this virus to precursor cells of the hematopoietic compartment (27).

MPSV derivatives containing the gene for neomycin resistance (Neo^r) are expressed in F9 cells but not in PCC4 or other embryonic cell lines (4, 9, 22). A second retroviral mutant (Neo^r PCC4 cell-passaged myeloproliferative sarcoma virus [PCMV]) with broader host range properties than MPSV was isolated from rare Neo^r colonies obtained after infection of PCC4 cells with Neo^r MPSV (4, 9). Analysis of the structural organization of the Neo^r PCMV genome shows two major divergences from that of Neo^r MPSV: (i) the U3 region of Neo^r PCMV contains only one copy of the 75-bp direct repeat sequences present in the U3 region of MPSV and (ii) a 250-bp deletion occurred in Neo^r PCMV at the junction between the *gag* and the Neo^r-encoding sequences (9).

In this study, we investigated the expression of recombinants between Neo^r MPSV and Neo^r PCMV. We showed that expansion of the host range of PCMV to PCC4 cells is conferred by rearrangements at the *gag-neo* junction. Furthermore, by using transient expression assays, we identified point mutations shared by the LTRs and leader sequences of MPSV and PCMV that are essential for virus expression in EC cells. One of these point mutations maps within the recognition site of an EC cell-specific DNA-binding factor.

(This work represents part of the doctoral thesis of E. Akgün [Faculty of Biology, University of Hamburg, Hamburg, Federal Republic of Germany].)

MATERIALS AND METHODS

Cell culture and virus titrations. NIH 3T3 and Psi-2 cells and EC cell lines F9 and PCC4 were grown in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum. F9 and PCC4 cells were grown on gelatin-coated plates. To obtain virus stocks, the retroviral constructs were transfected into packaging cell line Psi-2 (15). Virus supernatants were harvested from a pool of Neo^r colonies, filtered, and used to infect fibroblasts and EC cells. Briefly, cells were plated into 24-well plates at a density of 10^5 per well and inoculated 24 h later with 1 ml of diluted virus supernatant (3 wells per dilution). At 48 h postinfection, the virus supernatant was replaced by medium containing 0.2 mg of active G418 per ml. Medium was changed every 3 to 4 days. Neomycin-resistant colonies were counted after 12 to 14 days.

* Corresponding author.
† Present address: Georg-Speyer-Haus, Paul-Ehrlich-Strasse 42, 6000 Frankfurt 70, Federal Republic of Germany.

Downloaded from jvi.asm.org by on September 14, 2009

Plasmids. Recombinant plasmids containing retroviral LTRs upstream of the coding region of the gene for chloramphenicol acetyltransferase (CAT) were derived from pMMTV-CAT (17) by replacing the mouse mammary tumor virus LTR by the Mo-MuLV, Mo-MuSV, MPSV, or PCMV LTR from positions −418 to +31.

DNA transfections and CAT assays. NIH 3T3, F9, and PCC4 cells were plated at a density of 10^6/10-cm-diameter plate 6 h before transfection. A 10-μg sample of CAT-encoding plasmid DNA was used per plate. This DNA amount was within the linear range of CAT activity in all of the cell lines tested. Cell extracts were prepared 48 h posttransfection and tested for CAT activity as described by Gorman et al. (6).

Preparation of nuclear extracts. Nuclear extracts were prepared essentially as described by Gorski et al. (8). Briefly, cells (10^8 to 10^{10}) were washed with cold phosphate-buffered saline, detached from the tissue culture plates with a rubber policeman, and collected by centrifugation. The cell pellet was suspended in 30 ml of homogenization buffer (10 mM HEPES [*N*-2-hydroxyethylpiperazine-*N*'-2-ethanesulfonic acid] [pH 7.6], 25 mM KCl, 0.15 mM spermine, 0.5 mM spermidine, 1 mM EDTA, 2 M sucrose, 10% glycerol), homogenized, and layered on top of 10 ml of homogenization buffer in an SW27 Beckman ultra-clear tube. Nuclei were sedimented at 24,000 rpm for 30 min at 2°C in an SW27 rotor. Nuclear pellets (normally between 10^8 and 10^9 nuclei) were suspended in nuclear lysis buffer (10 mM HEPES [pH 7.6], 100 mM KCl, 3.0 mM $MgCl_2$, 0.1 mM EDTA, 1 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, 10% glycerol), homogenized, and diluted to 10 A_{260} units/ml. Nuclear proteins were extracted by adding 0.1 volume of 4 M $(NH_4)_2SO_4$ dropwise over a 30-min period. The suspension was cleared by centrifugation (35,000 rpm, 60 min in a 60 Ti rotor at 0°C). Proteins were concentrated by addition of 0.3 g of $(NH_4)_2SO_4$ per ml, followed by centrifugation at 35,000 rpm for 30 min in an SW27 rotor. The protein pellet was suspended in dialysis buffer (25 mM HEPES [pH 7.8], 40 mM KCl, 0.1 mM EDTA, 10% glycerol, 1 mM dithiothreitol, 1 mg of aprotinin per ml, 0.1 mM phenylmethylsulfonyl fluoride, 0.05% Nonidet P-40), dialyzed twice (2 h each time), cleared by centrifugation, and frozen in liquid nitrogen. Samples were kept in liquid nitrogen until use.

Gel retardation assay. Nuclear extracts were preincubated on ice with 1.8 μg of poly(dI-dC) in 10 mM HEPES (pH 7.9)–0.1 mM EDTA–20 mM KCl–4.0 mM $MgCl_2$–4.0 mM spermidine–2.0 mM dithiothreitol–17.5% glycerol in a total volume of 25 μl. After 15 min, 2 fmol of a ^{32}P-labeled DNA fragment (6,000 to 12,000 cpm) was added to the incubation mixture. After 30 min on ice, the reaction mixture was loaded onto an 8% polyacrylamide gel in 45 mM Tris-borate–1 mM EDTA and electrophoresed at 120 V for 4 h at room temperature.

RESULTS

Rearrangements of sequences at the *gag-neo* junction are essential for expansion of the host range of PCMV. The different host range properties of Neo^r MPSV and Neo^r PCMV prompted us to analyze the elements involved in expansion of the host range of PCMV. Recombinants between Neo^r MPSV and Neo^r PCMV were constructed and tested for the ability to confer neomycin resistance to PCC4 cells. An *XhoI-SphI* DNA fragment which includes the junction between the viral *gag* sequences and the gene for

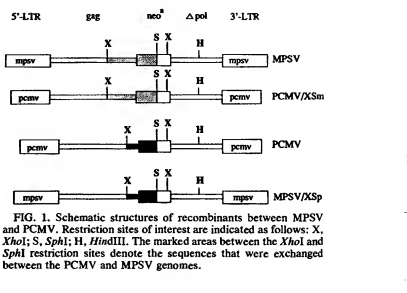


FIG. 1. Schematic structures of recombinants between MPSV and PCMV. Restriction sites of interest are indicated as follows: X, *Xho*I; S, *Sph*I; H, *Hin*dIII. The marked areas between the *Xho*I and *Sph*I restriction sites denote the sequences that were exchanged between the PCMV and MPSV genomes.

Neo^r was exchanged between the two genomes (Fig. 1). The resulting recombinants, MPSV/XSp and PCMV/XSm, and the parental plasmids were transfected into Psi-2 cells (15). Virus stocks were prepared from a pool of neomycin-resistant colonies and used to infect NIH 3T3 and PCC4 cells. Table 1 shows the result of such an experiment. As expected, infection of PCC4 cells with Neo^r MPSV resulted in a low number of G418-resistant colonies, while a 100-fold increase was observed after infection with Neo^r PCMV. Recombinant virus MPSV/XSp, which contains the XhoI-*Sph*I DNA fragment from Neo^r PCMV in an MPSV background, conferred neomycin resistance to PCC4 cells with a frequency that was at least 300-fold higher than that of wild-type Neo^r MPSV. The reciprocal recombinant, PCMV/XSm, showed a 100-fold reduction in titer compared with wild-type Neo^r PCMV. These results suggest that the rearrangements of sequences at the *gag-neo* junction in PCMV are crucial for expansion of the host range of the virus.

Sequence comparison between the *XhoI-SphI* DNA frag-

TABLE 1. Transfer of Neo^r to EC cells with Neo^r MPSV, Neo^r PCMV, and their recombinants[a]

Virus	Neo^r CFU on:		Ratio of Neo^r CFU on PCC4 cells/NIH 3T3 cells
	NIH 3T3 cells	PCC4 cells	
Neo^r MPSV	1.5×10^5	1.5	1.0×10^{-5}
	2.3×10^5	5.0	2.0×10^{-5}
Neo^r PCMV	4.0×10^5	8.3×10^2	2.1×10^{-3}
	1.9×10^5	3.8×10^2	2.0×10^{-3}
MPSV/XSp	1.2×10^6	1.3×10^4	1.1×10^{-2}
	3.0×10^5	2.0×10^3	6.6×10^{-3}
PCMV/XSm	4.5×10^5	13.2	3.0×10^{-5}
	1.1×10^5	0.0	

[a] The values shown represent the total number of neomycin-resistant colonies per milliliter of virus calculated from the number of clones obtained at the endpoint dilution of the virus supernatant. Infection and selection for Neo^r colonies were done as previously described (4). For each virus, titers were obtained from at least two independent experiments.

Downloaded from jvi.asm.org by on September 14, 2009

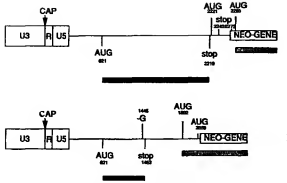


FIG. 2. Summary of the open reading frames in Neo^r MPSV and Neo^r PCMV. Black bars, reading frames for the *gag*-encoded protein; shaded bars, putative reading frames for neomycin phosphotransferase. A single-base-pair deletion occurs in the Neo^r PCMV genome at position +1445 (the numbering is that of Shinnick et al. [23]). As a consequence, a translation stop codon occurs shortly thereafter, generating a truncated p30 protein of only 62 amino acids.

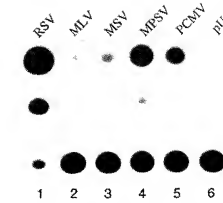


FIG. 3. Transcriptional activities of viral LTRs in EC cells. PCC4 cells were transfected with 10 μg of Rous sarcoma virus (RSV)-CAT (lane 1), MLV-CAT (lane 2), MSV-CAT (lane 3), MPSV-CAT (lane 4), PCMV-CAT (lane 5), or pUC-CAT (lane 6). Transfections and CAT assays were performed as described previously (9).

ments of Neo^r PCMV and Neo^r MPSV revealed a 247-bp deletion in the genome of Neo^r PCMV extending from position +2063 towards the coding region of the gene for neomycin resistance. This deletion removed the first 21 bp of the coding region of the *neo* gene, including the procaryotic translation initiation signals around position +2290 (Fig. 2). As a consequence, translation of the *neo* mRNA in PCMV depends on the presence of AUG codons within the *gag* region of the viral RNA. Two putative translation initiation signals in frame with the coding region of the gene for neomycin phosphotransferase are located at positions +1802 and +2039. Thus, replacement of the procaryotic translation initiation signals by eucaryotic ones may result in a more efficient translation of the neomycin phosphotransferase mRNA in mammalian cells and, therefore, in a larger number of geneticin-resistant PCC4 cells.

MPSV and PCMV LTRs are functional in EC cells. Efficient expression of Neo^r PCMV and MPSV/XSp in PCC4 cells suggested that both LTRs are functional in these cells. To test this hypothesis, the LTRs of MPSV and PCMV (−418 to +31) were joined to the coding region of the chloramphenicol acetyltransferase (CAT) gene. Similarly, the LTRs of Rous sarcoma virus, Mo-MuLV, and Mo-MuSV were inserted upstream of the CAT gene and served as positive (RSV-CAT) and negative (MLV-CAT and MSV-CAT) controls for CAT expression in PCC4 cells, respectively. These constructs were transfected into NIH 3T3 and PCC4 cells. At 48 h posttransfection, cell extracts were prepared and assayed for CAT activity as described by Gorman et al. (6). The levels of CAT expression in NIH 3T3 cells were similar for all recombinants (data not shown). As expected, expression of the CAT gene in plasmid RSV-CAT was high in PCC4 cells, while only background levels of CAT activity were detected after transfection of MLV-CAT or MSV-CAT (Fig. 3, lanes 1 to 3). Transfection of the MPSV-CAT and PCMV-CAT constructs into PCC4 cells resulted in conversion of similar amounts (20%) of chloramphenicol to its acetylated forms (Fig. 3, lanes 4 and 5). These results confirmed that the LTRs of MPSV and PCMV are functional in PCC4 cells and suggested that their transcriptional control regions share some common features not present in the LTR of Mo-MuLV or Mo-MuSV. Reexamination of the MPSV sequence (25) between positions −419 and −150 showed mutations similar to those previously defined for the PCMV LTR (9; Fig. 4). Single-base-pair deletions or substitutions were found to be shared by the LTRs of MPSV and PCMV at positions −381, −345, −326, and −166.

To investigate the significance of these point mutations in the activation of MPSV and PCMV, a 5′ deletion analysis of the MPSV LTR was performed. Convenient restriction sites were used to generate truncated forms of the viral LTR.

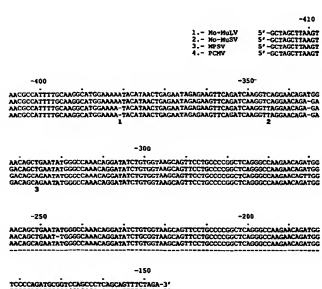


FIG. 4. DNA sequence comparison of the LTRs of PCMV (9), MPSV (25), Mo-MuSV (2), and Mo-MuLV (23). Only sequences between positions −419 and −147 are shown. The numbers beneath the sequences indicate point mutations common to the LTRs of MPSV and PCMV.

Downloaded from jvi.asm.org by on September 14, 2009

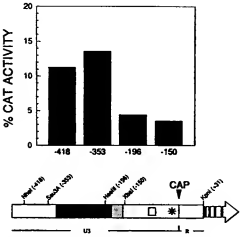


FIG. 5. Functional analysis of the enhancer region of MPSV in F9 cells. Restriction sites were used to generate a sequential 5' deletion of the MPSV LTR. The truncated LTRs were joined to the coding region of the CAT gene and transfered into F9 cells. CAT activity was measured 48 h posttransfection. Dark grey areas: 75-bp direct repeat; light grey region, G+C-rich region; arrow, CAT gene; *, TATA box; □, CAAT box.

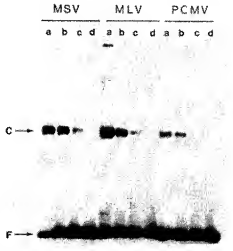


FIG. 6. The Mo-MuLV and Mo-MuSV LTRs are targets for DNA-binding proteins present in nuclear extracts obtained from undifferentiated PCC4 cells. The *Sau*3A-*Eco*RV fragments obtained from the Mo-MuLV, Mo-MuSV, and PCMV LTRs were inserted into the polylinker region of pUC19, excised with *Eco*RI and *Hin*dIII, labeled at the 3' end, and incubated with 4.75 (lanes a), 2.38 (lanes b), or 1.35 (lanes c) μg of crude nuclear extract from PCC4 cells or with 10 μg of bovine serum albumin (lanes d). The free (F) and bound (C) DNAs were resolved in an 8% polyacrylamide gel.

Thus, in MPSV/S-CAT, MPSV/H-CAT, and MPSV/X-CAT, sequences upstream of position −354 (*Sau*3A), −196 (*Hae*III), and −150 (*Xba*I), respectively, were deleted. These constructs and the parental MPSV-CAT recombinant were transfected into F9 cells. Cell extracts were prepared 48 h posttransfection and tested for CAT activity. MPSV/S-CAT showed levels of CAT activity similar to those of the parental MPSV-CAT construct, while a threefold reduction in CAT activity was observed after transfection of MPSV/H-CAT or MPSV/X-CAT (Fig. 5). These results suggested that sequences between positions −354 and −196 are involved in activation of the MPSV LTR in EC cells. This 158-bp fragment contains two point mutations (−345 and −326) which are conserved between the LTRs of MPSV and PCMV and absent from the Mo-MuSV and Mu-MuLV LTRs (Fig. 4).

Binding of cellular factors to sequences between −354 and −306 correlates with retroviral expression in EC cells. To elucidate possible mechanisms by which these mutations affect viral expression in EC cells, nuclear extracts were prepared from undifferentiated PCC4 cells and tested by the gel retardation assay for sequence-specific DNA-binding activities. The 47-bp, ^{32}P-labeled *Sau*3A-*Eco*RV fragments (−354 to −306) obtained from the Mo-MuLV, Mo-MuSV, and MPSV-PCMV LTRs were incubated with increasing amounts of nuclear extracts and applied to an 8% polyacrylamide gel (Fig. 6). A single specific complex of identical mobility was formed with each fragment. Similar amounts of Mo-MuSV and Mo-MuLV DNAs were incorporated into a protein-DNA complex and thus migrated slower than free DNA. In contrast, the MPSV-PCMV fragment was bound to a much lesser extent. This nuclear factor was named EC cell factor I (ECF-I).

The binding site of ECF-I was localized between nucleotides −355 and −341 on the viral LTR by DNase I footprinting analysis (data not shown). To confirm this observation, oligonucleotides spanning sequences between nucleotides −363 and −335 were synthesized and used in gel retardation assays. As expected, the oligonucleotide containing the wild-type Mo-MuLV sequences (oMLV) was bound more efficiently than the PCMV-containing oligonucleotide (oPCMV) (Fig. 7). Scanning analysis of the bands in the shifting gel showed a two- to threefold difference in the amount of bound DNA between the oMLV and oPCMV oligonucleotides. The specificity of ECF-I for the Mo-MuLV sequences was confirmed by competition experiments in which the oMLV and oPCMV oligonucleotides were concatemerized and used as tetramers. Increasing molar concentrations of the tetramers were added to the PCC4 nuclear extracts before addition of radiolabeled oMLV. Similar to

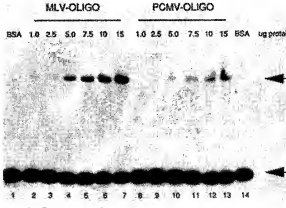


FIG. 7. Nuclear factor binding to oligonucleotides oMLV and oPCMV. End-labeled oligonucleotide oMLV or oPCMV (1 fmol) was incubated with increasing amounts (given in micrograms at the tops of lanes 2 to 13) of nuclear extracts prepared from undifferentiated PCC4 cells as described in Materials and Methods. The oligonucleotides in lanes 1 and 14 were incubated with 15 μg of bovine serum albumin (BSA). The locations of free (F) and bound (C) DNAs are shown.

Downloaded from jvi.asm.org by on September 14, 2009

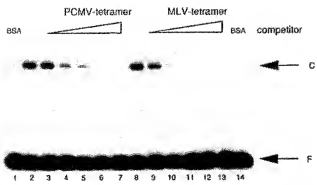


FIG. 8. Specificity of binding to oligonucleotide oMLV. Nuclear extracts from PCC4 cells (10 μg) were incubated with a 12-fold (lanes 3 and 9), 60-fold (lanes 4 and 10), 120-fold (lanes 5 and 11), 480-fold (lanes 6 and 12), or 1,200-fold (lanes 7 and 13) molar excess of an oMLV (lanes 9 to 13) or oPCMV (lanes 3 to 7) tetramer before addition of a radiolabeled oMLV monomer (1 fmol) to the incubation mixture. In lanes 2 and 8, the oMLV oligonucleotide was incubated with no specific competitor, whereas in lanes 1 and 14, the nuclear extract was replaced with bovine serum albumin (BSA; 10 μg). The locations of free (F) and bound (C) DNAs are shown.

the binding results described above, a 2.5-fold molar excess of the PCMV tetramer was required for competition like that of the oMLV tetramer (Fig. 8), whereas a 2,400-fold excess of a tetramer containing multiple point mutations in the ECF-I recognition site did not affect binding of ECF-I to the oMLV oligonucleotide (data not shown).

DISCUSSION

Isolation of retroviral host range mutants for EC cells allowed us to define individual mutations required for expression of retroviruses in EC cells. For this analysis, we used two retroviral mutants with distinct host range properties. Neo^r MPSV confers G418 resistance at high efficiency to F9 cells but at low efficiency to PCC4 cells. In contrast, Neo^r PCMV induces G418 resistance in both cell lines at high efficiency. Recombinants between Neo^r MPSV and Neo^r PCMV showed that the expanded host range property of Neo^r PCMV was conferred by a novel junction between the viral *gag* and *neo* sequences as a consequence of a 250-bp deletion that occurred during passage of Neo^r MPSV in PCC4 cells. We favor the hypothesis that this new configuration provides a better context of sequences for efficient translation of Neo^r-encoding mRNA in mammalian cells. Furthermore, deletion of all *gag* sequences in Neo^r MPSV resulted in no significant increase in the number of neomycin-resistant PCC4 colonies obtained (33), suggesting that deletion of the procaryotic sequences was the crucial event in expansion of the host range of MPSV. Similar alterations have been made by genetic manipulation of eucaryotic plasmid vectors containing the *neo* gene. Replacement of the procaryotic translation initiation signals intrinsic to the Neo^r-encoding mRNA by a synthetic oligonucleotide optimized to direct efficient translation of Neo^r-encoding mRNA in eucaryotic cells contributed to the large increase in the number of G418-resistant colonies obtained (30).

Rearrangement of the *gag-neo* junction is not the only structural feature that allows expansion of the host range of MPSV to PCC4 cells. Replacement of MPSV sequences between positions +31 and +284 by a similar region obtained from retroviral mutant B2 also resulted in an increased frequency of G418-resistant PCC4 cells (33). Most significant, however, is the observation that the LTRs and 5′ leader sequences of MPSV and PCMV are compatible with retrovirus replication and expression in EC cells.

A DNA sequence comparison between the LTRs of Mo-MuLV, Mo-MuSV clone m1, and MPSV revealed five base pair changes within the LTR of MPSV that are not present in the LTR of Mo-MuLV or Mo-MuSV. These five point mutations are common to MPSV and PCMV and thus must be involved in transcriptional activation of these viruses in EC cells. Four of the point mutations are located within the U3 region (−381, −345, −326, and −166), while the fifth is located within the R region (+44) of the viral genome. We do not know whether all five point mutations are required for viralactivity in EC cells. However, the CAT constructs used to estimate the expression level of the MPSV and PCMV LTRs in EC cells contained viral sequences up to position +31 and thus excluded the point mutation at position +44. These constructs were functional in EC cells, suggesting that the point mutation at position +44 is not essential but may be required for efficient viral expression in EC cells. Furthermore, functional analysis of the 5′ deletion mutants showed that the base pair deletion at position −381 is also dispensable for retroviral expression in EC cells. Our results suggest that expression of MPSV and PCMV in EC cells correlates with the presence of three point mutations (−345, −326, and −166) within the enhancer region of the viral genome.

This observation was at first intriguing, since others have reported that a region 3′ from the LTR is also involved in restricting viral expression in EC cells (1, 3, 13, 14). The inhibitory region has been mapped to a 29-bp element between positions +147 and +175 in the Mo-MuLV genome (13, 14). Close examination of the nucleotide sequence of Mo-MuSV (2) revealed two point mutations within this region at positions +164 and +165 which are conserved in MPSV and PCMV. Moreover, Feuer et al. (3) have shown that leader sequences containing both base pair alterations do not affect transient expression from 5′-positioned promoters in F9 cells. Thus, the 5′ leader region of Mo-MuSV is compatible with retroviral expression in EC cells. This observation explains our success in obtaining retroviral host range mutants for EC cells. MPSV was isolated as a spontaneous mutant of Mo-MuSV, an already mutated retrovirus with leader sequences permissive for expression in EC cells. Furthermore, the U3 region of Mo-MuSV differs from that of Mo-MuLV by several base pair deletions or substitutions that may be required for efficient expression in EC cells. Thus, Mo-MuSV may be considered as partially restricted in EC cells, since the only restraint to Mo-MuSV expression in EC cells is the lack of transcriptional activity of the enhancer region. The additional mutations observed in the U3 region of MPSV must then contribute to the enhancement of viral transcription in EC cells.

Since enhancement of transcription is mediated by cellular DNA-binding factors, the point mutations in the enhancer regions of MPSV and PCMV may either impair binding of repressor molecules to the viral LTR or provide new binding sites for transcription activators. We detected a DNA-binding factor (ECF-I) which binds strongly to the LTRs of Mo-MuLV and Mo-MuSV but only weakly to the LTRs of MPSV and PCMV. A DNA-binding factor (embryonic LTR-

Downloaded from jvi.asm.org by on September 14, 2009

binding protein ELP) which recognizes similar sequences has been described by Tsukiyama et al. (31). The sequences protected by ELP from DNase I digestion are similar to the footprint pattern of ECF-I (34). We believe that ELP and ECF-I are identical or very similar factors. Since ELP–ECF-I activity is detected only in undifferentiated EC cells (31, 34) and binding of ELP–ECF-I correlates with lack of retrovirus expression in these cells, one could speculate that ELP–ECF-I is one of the factors responsible for the block of retroviral expression in EC cells. Furthermore, the recognition site of ELP–ECF-I includes the motif 5'-CTGACC-3', which is also present within the recognition site of PEA2, a nuclear factor that binds to the polyomavirus α-enhancer region (21). Mutations at positions 3 and 5 in the consensus hexamer sequence destroy binding of ECF-I–ELP and PEA2, respectively. Both residues are directly involved in the interaction of ELP–ECF-I with its cognate recognition site (31), suggesting that ELP–ECF-I and PEA2 are similar or related factors. Interestingly, PEA2 has been shown to be a repressor of enhancer activity in EC cells (32). All of this circumstantial evidence suggests that ECF-I is a negatively acting factor. However, neither introduction of a point mutation at position −345 or −326 into the Mo-MuLV enhancer region nor a combination of both results in a functional LTR in EC cells. Similarly, reversion of the point mutation at −345 in the PCMV enhancer to wild-type sequences does not impair retrovirus expression in EC cells (34). Thus, additional or other alterations are required for activation of retroviral enhancers in EC cells. The mutations present in the U3 region of MPSV and PCMV will help to resolve this issue.

ACKNOWLEDGMENTS

We thank C. Stocking, J. Nowock, and W. Ostertag for critical comments on the manuscript.

This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Stiftung Volkswagenwerk. The Heinrich-Pette-Institut is financially supported by the Freie und Hansestadt Hamburg and the Bundesministerium für Jugend, Familie, Frauen, und Gesundheit.

REFERENCES

1. **Barklis, E., R. C. Mulligan, and R. Jaenisch.** 1986. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. Cell 47:391–399.
2. **Dhar, R., W. L. McClements, L. W. Enquist, and F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA 77:3937–3941.
3. **Feuer, G., M. Taketo, R. Hanecak, and H. Fan.** 1989. Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. J. Virol. 63:2317–2324.
4. **Franz, T., F. Hilberg, B. Seliger, C. Stocking, and W. Ostertag.** 1986. Retroviral mutants efficiently expressed in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 83:3292–3296.
5. **Gautsch, J. W., and M. C. Wilson.** 1983. Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature (London) 301:32–37.
6. **Gorman, C., L. Moffat, and B. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2:1044–1051.
7. **Gorman, C. M., P. W. J. Rigby, and D. P. Lane.** 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell 42:519–526.
8. **Gorski, K., M. Carneiro, and U. Schibler.** 1986. Tissue-specific in vitro transcription from the mouse albumin promoter. Cell 47:767–776.
9. **Hilberg, F., C. Stocking, W. Ostertag, and M. Grez.** 1987. Functional analysis of a retroviral host-range mutant: altered long terminal repeat sequences allow expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 84:5232–5236.
10. **Hogan, B.** 1981. From embryo to teratocarcinoma in tissue culture. Nature (London) 292:111–112.
11. **Klein, B., C. Le Bousse, B. Fagg, F. Smajda-Joffe, K. Vehmeyer, K. J. Mori, C. Jasmin, and W. Ostertag.** 1981. Effects of myeloproliferative sarcoma virus on the pluripotential stem cells and granulocyte precursor cell populations of DBA/2 mice. JNCI 66:935–940.
12. **Linney, E., B. Davis, J. Overhauser, E. Chao, and H. Fan.** 1984. Non-function of a Moloney murine leukemia virus regulatory sequence in F9 embryonal carcinoma cells. Nature (London) 308:470–472.
13. **Loh, T. P., L. L. Sievert, and R. W. Scott.** 1987. Proviral sequences that restrict retroviral expression in mouse embryonal carcinoma cells. Mol. Cell. Biol. 7:3775–3784.
14. **Loh, T. P., L. L. Sievert, and R. W. Scott.** 1988. Negative regulation of retrovirus expression in embryonal carcinoma cells mediated by an intragenic domain. J. Virol. 62:4086–4095.
15. **Mann, R., R. C. Mulligan, and D. Baltimore.** 1983. Construction of a retrovirus packaging mutant and its use to produce helper free defective retroviruses. Cell 33:153–159.
16. **Martin, G. R.** 1980. Teratocarcinomas and mammalian embryogenesis. Science 209:768–776.
17. **Miksicek, R., U. Borgmeyer, and J. Nowock.** 1987. Interaction of the TGGCA-binding protein with upstream sequences is required for efficient transcription of mouse mammary tumor virus. EMBO J. 6:1355–1360.
18. **Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara.** 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell 32:1105–1113.
19. **Ostertag, W., K. Vehmeyer, B. Fagg, B. Klein, C. Jasmin, and H. Eisen.** 1980. Myeloproliferative virus, a cloned murine sarcoma virus with spleen focus-forming properties in adult mice. J. Virol. 33:573–582.
20. **Perles, J., E. Alves-Cardoso, M. Canivet, M. C. Devons-Guillemin, and J. Lasneret.** 1977. Lack of multiplication of ecotropic murine C-type viruses in mouse teratocarcinoma primitive cells. J. Natl. Cancer Inst. 59:463–465.
21. **Piette, J., and M. Yaniv.** 1987. Two different factors bind to the α-domain of the polyoma virus enhancer, one of which also interacts with the SV40 and c-fos enhancers. EMBO J. 6:1331–1337.
22. **Seliger, B., R. Kollek, C. Stocking, T. Franz, and W. Ostertag.** 1986. Viral transfer, transcription and rescue of a selectable myeloproliferative sarcoma virus in embryonal cell lines: expression of the *mos* oncogene. Mol. Cell. Biol. 6:286–293.
23. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) 239:543–548.
24. **Speers, W. C., J. W. Gautsch, and F. J. Dixon.** 1980. Silent infection of murine embryonal carcinoma cells by Moloney murine leukemia virus. Virology 105:241–244.
25. **Stacey, A., C. Arbuthnott, R. Kollek, L. Coggins, and W. Ostertag.** 1984. Comparison of myeloproliferative sarcoma virus with Moloney murine sarcoma virus variants by nucleotide sequencing and heteroduplex analysis. J. Virol. 50:725–732.
26. **Stewart, C. L., H. Stuhlmann, D. Jähner, and R. Jaenisch.** 1982. De novo methylation, expression and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 79:4098–4102.
27. **Stocking, C., R. Kollek, U. Bergholz, and W. Ostertag.** 1985. Long terminal repeat sequences impart hemopoietic transformation properties to the myeloproliferative sarcoma virus. Proc. Natl. Acad. Sci. USA 82:5746–5750.
28. **Strickland, S.** 1981. Mouse teratocarcinoma cells: prospects for the study of embryogenesis and neoplasia. Cell 24:277–278.

Downloaded from jvi.asm.org by on September 14, 2009

29. **Teich, N. M., R. A. Weiss, G. R. Martin, and D. R. Lowy.** 1977. Virus infection of murine teratocarcinoma stem cell lines. Cell 12:973–982.
30. **Thomas, K. R., and M. R. Capecchi.** 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 51:503–512.
31. **Tsukiyama, T., O. Niwa, and K. Yokoro.** 1989. Mechanism of suppression of the long terminal repeat of Moloney leukemia virus in mouse embryonal carcinoma cells. Mol. Cell. Biol. 9:4670–4676.
32. **Wasylyk, B., J. L. Imler, B. Chatton, C. Schatz, and C. Wasylyk.** 1988. Negative and positive factors determine the activity of the polyoma virus enhancer α domain in undifferentiated and differentiated cell types. Proc. Natl. Acad. Sci. USA 85:7952–7956.
33. **Weiher, H., E. Barklis, W. Ostertag, and R. Jaenisch.** 1987. Two distinct sequence elements mediate retroviral gene expression in embryonal carcinoma cells. J. Virol. 61:2742–2746.
34. **Ziegler, M., and M. Grez.** Unpublished data.

Downloaded from jvi.asm.org by on September 14, 2009

TAB 3

This material may be protected by Copyright law (Title 17 U.S. Code)

Creation of genomic methylation patterns

Timothy H. Bestor[1] & Benjamin Tycko[2]

There are two biological properties of genomic methylation patterns that can be regarded as established. First, methylation of 5'-CpG-3' dinucleotides within promoters represses transcription, often to undetectable levels. Second, in most cases methylation patterns are subject to clonal inheritance. These properties suit methylation patterns for a number of biological roles, although none of the current hypotheses can be regarded as proved or disproved. One hypothesis suggests that the activity of parasitic sequence elements is repressed by selective methylation. Features of invasive sequences that might allow their identification and inactivation are discussed in terms of the genome defense hypothesis. Identification of the cues that direct *de novo* methylation may reveal the biological role (or roles) of genomic methylation patterns.

[1]*Department of Genetics and Development*
[2]*Department of Pathology, Columbia University, College of Physicians and Surgeons, 701 W. 168th Street New York, New York 10032, USA.*

Correspondence should be addressed to T.H.B.

The mammalian genome is ornamented with ~3×10^7 methyl groups, all at the 5 position of cytosine (m^5C) and most at 5'-CpG-3' dinucleotides. Methylation patterns increase the information content of the genome[1] and are transmitted by clonal inheritance[2]; methylation of CpG sites within promoters represses transcription[3]. This natural modification is also dangerous: m^5C is the major endogenous mutagen (deamination results in C→T transition mutations at CpG sites, which account for about one-third of all mutations in humans)[4], and tumour suppressor genes are frequently inactivated by ectopic *de novo* methylation of promoter regions[5,6]. However, there must be benefits that yield a net selective advantage. This is shown by the retention of cytosine methylation by virtually all organisms with genomes >5×10^8 basepairs[7], and by the demonstration that perturbations of methylation patterns are lethal to mouse embryos and to differentiated cells[8,9]. While methylation patterns clearly provide some essential function, the nature of that function or functions is still enigmatic.

There are at present two salient hypothetical roles of genomic methylation patterns. The first hypothesis (put forward twenty years ago) suggests that programmed demethylation and *de novo* methylation play a direct role in gene control during development[10,11]. Methylation patterns would be established during gametogenesis or early development, and regulatory factors would mediate the removal of methyl groups from promoters to allow the expression of tissue-specific genes at the appropriate stage of differentiation. While supported by a large body of indirect and correlative evidence, a definitive test of the causality of cytosine methylation in developmental gene control has been elusive. By the same token, a developmental role for cytosine methylation has not yet been disproved. It seems likely that methylation patterns might reinforce the heritability of states of gene expression mediated by chromatin proteins analogous to the Polycomb and trithorax group of proteins from *Drosophila*[12,13].

The second hypothesis suggests that cytosine methylation is part of a genome defense system which inactivates parasitic sequences such as transposable elements and proviral DNA[4,7]. Most of these elements are in fact methylated and transcriptionally inert in the genome of mammals, flowering plants, and those fungi whose genomes contain m^5C. Treatment of cultured cells or mice with the demethylating drug 5-azacytidine can activate silent retroviruses and endogenous genes that have been silenced by ectopic *de novo* methylation of regulatory regions[14]. It is striking that *Drosophila*, whose DNA lacks m^5C, suffers far larger numbers of insertion mutations than do animals whose genomes are methylated[15]. These observations, with support from evolutionary considerations[7], strongly suggest that cytosine methylation is part of a genomic host defense system that suppresses the transcription of parasitic sequence elements. The selective advantage of such a defensive system is obvious, given that a sizable fraction of the genome represents parasitic sequences that are invisible to the immune system and which might inflict intolerable mutational or cytotoxic damage if allowed to proliferate unchecked.

These hypothetical roles of genomic methylation in development or host defense place fundamentally different requirements on the DNA methylating system. The developmental role requires that methylation patterns be established as part of the developmental program via conventional sequence recognition mediated by sequence-specific DNA methyltransferases or specificity factors that interact with the ubiquitous DNA methyltransferase. The host-defense function requires a completely different

44/264

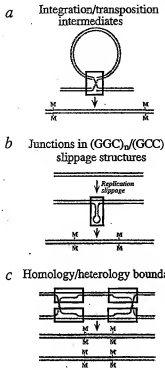


Fig. 1 Structure-dependent *de novo* methylation. Complementary sequences are depicted in identical colors, and structural features that are proposed to trigger silencing are boxed. M: 5-methylcytosine. *a*, A retrovirus integration intermediate showing the stage at which free 3' ends of the proviral DNA have been linked to 5' ends of chromosomal DNA. Most transposition and retroviral integration events utilize a common mechanism[19], and integrases and transposases share sequence or structural similarities. Recognition of integration/transposition intermediates by the silencing system could protect the genome from the proliferation of parasitic sequences. *b*, formation of three-way junctions in replication slippage structures at (GGC)/(GCC) repeats. Nearly all triplet repeat expansions at (GGC)/(GCC) sequences trigger *de novo* methylation in and around the repeats, and it is suggested that the three-way junction is the stimulus. Silencing of the *FMR1* gene in Fragile X syndrome is suggested to occur via this pathway. *c*, Detection of invasive sequences via homology/heterology boundaries during strand exchange[22]. Recognition and methylation of the boxed regions would result in reciprocal silencing of repeated sequences that reside in different sequence contexts.

source of specificity; *de novo* methylation must be cued by dynamic structural features unique to parasitic sequences in the act of invading the genome, or by events such as transient pairing interactions of repetitive elements. In other words, the developmental role requires that methylation be *directed* to specific sequences as part of the developmental programme, while the host defense role requires that methylation be *contingent* on the interaction of features characteristic of parasitic sequence elements with the DNA methylating system. Therefore it should be possible to deduce the role of methylation patterns from the nature of the biochemical events by which they were established.

Sequences that attract or repel *de novo* methylation in transfected embryonic cells have been identified[16–18], but many sequences (both foreign and endogenous) can be stably propagated in either the methylated or unmethylated state, and *de novo* methylation of such elements must therefore be dependent on features other than simple sequence recognition. The remainder of this article discusses possible mechanisms by which such sequences might be recognized and silenced.

Invasive DNA *in flagrante*

The host defense hypothesis requires that the silencing apparatus recognize and inactivate parasitic sequence elements. Nearly all transposition and viral integration intermediates share certain structural features[19], and some satellite DNA is thought to undergo amplification by extrachromosomal rolling circle replication followed by insertion of the array into the genome[20]. Recognition and *de novo* methylation of CpG sites in and around features characteristic of integration reactions would insure the inactivation of the invasive element immediately upon its integration. DNA methyltransferase may have an intrinsic ability to recognize integration intermediates that are characteristic of the above integration events. The *de novo* sequence specificity of mammalian DNA methyltransferase is strongly dependent on alternative secondary structures in DNA; four-way junctions in cruciform structures formed by inverted repeats in supercoiled plasmids are especially favored targets[21], as are secondary structures in artificial oligonucleotide substrates[22]. This biochemical property suggests that invasive sequences might be targeted for *de novo* methylation because of their presentation of alternative secondary structures during integration (Fig 1*a*).

Three-way junctions share structural features with four-way junctions, and enzymes that recognize four-way junctions (such as T4 endonuclease VII (ref. 23) and T7 endonuclease I (ref. 24)) recognize three-way junctions as well. The nature of the interaction of the methylating system with three-way junctions is of special interest because it may be involved in the aetiology of human diseases associated with GGC/GCC triplet repeat expansions. Slippage of replication intermediates can result in extrusion of a segment of GGC/GCC repeats, which form stable three-way junctions despite a lack of perfect complementarity in the extruded segment[22,25](Fig. 1*b*). The preference of DNA methyltransferase for junctions may result in *de novo* methylation of sequences in slippage structures. Diseases that result from GGC/GCC triplet repeat expansions may be thought of as autoimmune disorders, in which a host defense system attacks an innocuous alteration of an endogenous gene because of its incidental similarity to a parasitic sequence element in the act of integration.

Cytosine methylation and pairing interactions

A common characteristic of invasive sequences is their presence in multiple copies, and it has recently become known that repeated sequences can interact so as to trigger their mutual silencing. Fungi and flowering plants have diverse and highly effective means of silencing repeated sequences, and cytosine methylation is associated with silencing in nearly all cases. RIP (repeat-induced point mutation) imposes methylation, silencing, and large numbers of C→T transition mutations on repeated sequences during the sexual phase of the fungus *Neurospora crassa*[26], while MIP (methylation induced premeiotically) inactivates and methylates repeats during the sexual cycle of the fungus *Ascobolus immersus*[27]. Flowering plants can also silence and methylate repeated sequences via RIGS (repeat-induced gene silencing)[28, 29], the efficiency is such that multicopy transgenes frequently cause mutual silencing of themselves and of homologous resident sequences. The process can be remarkably efficient; unlinked transgene sequences as small as 300 bp can identify and inactivate each other in a genome of $>10^9$ bp.

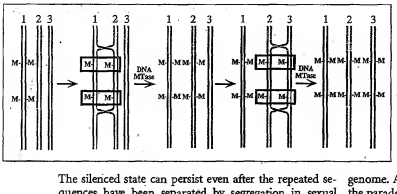


Fig. 2 Unidirectional transfer of epigenetic information via paired hemimethylated intermediates. Strand exchange between methylated and unmethylated repeated sequences presents hemimethylated sites, which are the preferred substrate of DNA methyltransferase; note that the maintenance activity of DNA methyltransferase leads to *de novo* methylation under these conditions. A common methylation pattern could propagate through an array or network of repeats by this mechanism, and methylation patterns could be transferred between alleles at loci that bear allele-specific methylation patterns.

The silenced state can persist even after the repeated sequences have been separated by segregation in sexual crosses[28]. Transgene silencing of this type is emerging as a barrier to the improvement of commercially-important plant species[30,31]. It is difficult to imagine a selective advantage for RIP, MIP, or RIGS other than host defense against the proliferation of parasitic sequence elements.

Transient pairing interactions are likely to be involved in the imposition of silencing and methylation on repeats; two mechanisms can be envisioned. First, strand exchange between repeats at non-allelic positions will present an abrupt loss of homology at the junction with flanking sequences (Fig. 1*c*), and silencing may be provoked by factors that recognize characteristic structural features at the homology/heterology boundary. This idea was developed by Signer and colleagues[32] to explain copy number-dependent transgene silencing in *Arabidopsis*. Second, methylation might propagate through a network of repeated sequences once one or more copies are methylated (Fig. 2). Strand exchange between methylated and unmethylated repeated sequences creates hemimethylated intermediates, which as mentioned previously, are strongly preferred substrates of DNA methyltransferase[12]; the preference for hemimethylated substrates normally contributes to maintenance methylation, but under these conditions *de novo* methylation is the result. Strand exchange reactions (which are thought to occur frequently as part of the double strand break repair pathway[33]) (Fig. 2) could cause a common methylation pattern to propagate through a network of repeats[28,30], and the probability of pairing with a methylated repeat will increase in proportion to the number of methylated repeats. A large number of methylated resident repeats will therefore increase the probability that a new, unmethylated copy will be methylated soon after its insertion into the genome. A large burden of parasitic sequences may have the paradoxical effect of repressing their own activity and that of homologous invasive sequences. Natural selection may actually favor the retention of large numbers of inactive parasitic sequences for this reason.

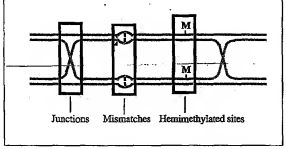


Fig. 3 Vulnerability of heteroduplex recombination intermediates to *de novo* methylation. Crossing-over during meiosis presents DNA methyltransferase with highly preferred targets: four-way junctions, mismatches and hemimethylated sites. Dysregulated *de novo* methylation appears to be prevented by sharp down-regulation of DNA methyltransferase during the pachytene stage of meiosis; the mechanism involves the production of a larger, non-translated DNA methyltransferase mRNA[46].

Many endogenous genes are present in multiple copies which escape repeat-dependent gene silencing. In *Neurospora*, which seems to be especially aggressive in its response towards repeated sequences, the efficiency of RIP is greater when the repeats are in close proximity[26]. This argues for special protective mechanisms that shield repeated cellular genes from repeat-induced silencing. A complete intolerance of repeated sequences would also put severe constraints on the evolution of new functions, which depends on the duplication and divergence of existing genes. However, the factors that control sensitivity to repeat-dependent silencing are poorly understood[34].

There is no clear experimental evidence that repeat-dependent silencing or methylation operates in mammals, as is the case in fungi and plants. However, no objective search for repeat-dependent silencing in mammals seems to have been conducted. Circumstantial evidence suggests that it may occur[35]. Much of the m^{2}C in the mammalian genome is found in repetitive DNA, and most of the repeated and potentially transposable elements in the genome are methylated and transcriptionally inert through most of development[36,37]. Retroviral vectors that transduce reporter genes or therapeutic agents have been observed to undergo methylation silencing after variable periods of expression in animals[38]. Susceptibility to *de novo* methylation and silencing has limited the usefulness of retroviral vectors in the construction of transgenic mouse lines, and it is probable that silencing phenomena may emerge as barriers to long-term somatic gene therapy in humans. Successful gene transfer may require the development of delivery vectors that evade the silencing response. It is also possible that the development of therapeutic agents that perform selective methylation of deleterious sequence elements (such as HIV-1 proviral DNA) will activate an existing host defense system and therefore serve as a sort of nuclear vaccine.

Maintenance of allele-specific methylation patterns

Mammalian genes whose expression is dependent on the sex of the contributing parent are said to be imprinted; such genes bear allele-specific methylation patterns that are necessary for maintenance of the imprinted state[9]. Short repetitive sequences are characteristic of imprinted genes[39,40]; these short repeats may form alternative secondary structures that trigger *de novo* methylation in the germ line or during early postfertilization development. A parent-of-origin effect on gene expression results if there

are differences in the types of DNA or chromatin structures formed, or if the response to the structures differs, in the male and female germlines. In this way a nuclear host defense system may give rise to allele-specific methylation patterns and to the genetic phenomenon of imprinting[41], and could also account for documented gamete-of-origin effects on methylation of repetitive DNA[37]. It should be noted that several mutations that result from insertion of retroviruses of the intracisternal A particle (IAP) type confer imprinted behavior on genes that are not imprinted in the wild type[42,43].

Alleles at imprinted loci are asymmetrically methylated, and the mechanism shown in Fig. 2 will tend to convert asymmetrical allelic methylation patterns towards the more heavily methylated pattern. Certain pathological human conditions show abnormalities in the functional imprinting of particular chromosomal regions which might arise via this type of interchromosomal transfer of epigenetic information. For example, Wilms' tumours frequently show conversion to a uniparental (bipaternal) methylation and expression pattern at imprinted loci in the *H19/IGF2* region on chromosome 11p15.5, which could result from the local transfer of methylation patterns from the paternal chromosome to the less heavily-methylated maternal chromosome[44]. It is important to note, however, that a conversion to a symmetrical uniparental allelic methylation pattern at imprinted loci could also come about by an alternative mechanism that involves deletion or mutations of a *cis*-acting DNA element (an 'imprinting centre'). This has been implicated in certain kindreds that show disruption of imprinting patterns within the chromosome 15 Prader Willi/Angelman syndrome region[45].

Vulnerability of meiotic recombination intermediates

Allelic heteroduplex recombination intermediates present several features that should make them very vulnerable to *de novo* methylation (Fig. 3). First, allelic methylation differences create hemimethylated sites in the heteroduplexes, which provoke *de novo* methylation via the maintenance activity of DNA methyltransferase (see Fig. 2). Second, Holliday structures and four-way junctions are necessarily present; as described earlier, these structures appear to be favored targets of DNA methyltransferase. Third, mismatches in the vicinity of CpG dinucleotides greatly favor *de novo* methylation, presumably by lowering the energetic barrier associated with eversion of the target cytosine during the transmethylation reaction[46,47]. Presentation of these vulnerable sites might create a "methylation ratchet" in which methylation levels increase in an unregulated fashion with each meiotic cycle.

Examination of purified germ cells from male mice showed that both DNA methyltransferase protein and the 5.2-kb DNA methyltransferase mRNA found in all proliferating cell types was present in all germ cell fractions, except for pachytene spermatocytes. These spermatocytes showed an absence of DNA methyltransferase protein and contained a 6.2-kb RNA that was not associated with polyribosomes. It is the pachytene stage of meiosis I where most crossing-over occurs. The ubiquitous 5.2-kb DNA methyltransferase mRNA and DNA methyltransferase protein reappeared at the conclusion of the crossing-over phase of meiosis[48]. These findings suggest that meiotic recombination intermediates are protected from *de novo* methylation through down-regulation of DNA methyltransferase via a novel post-transcriptional mechanism that involves the production of a larger, non-translated RNA transcript.

The function of cytosine methylation

The fact that cytosine methylation can increase the information content of DNA has tempted many to attribute diverse roles to methylation patterns. Cytosine methylation has been proposed to reduce the effective size of the genome by masking non-regulatory regions in large-genome eukaryotes[7,49], and central roles in DNA repair[22,50] and replication[51] have also been mooted. None of the hypothetical functions of cytosine methylation (and this includes the developmental and host-defense functions) has the support of compelling experimental evidence, and all, some, or none of the hypotheses may be valid. At this time it seems that the true function of cytosine methylation will be understood only when we learn how the cell selects specific sequences for covalent modification.

Aknowledgements

E. Selker and M. Turker made helpful comments on the manuscript. Supported by NIH grants GM00616 and CA60610 (to T. H. B.) and NIH grant CA60765 and American Cancer Society grant JFRA482 (to B.T.).

Received 27 November 1995; accepted 3 January 1996.

1. Reinisch, K. M., Chen, L., Verdine , G.L. & Lipscomb, W.N. (1995) The crystal structure of HaeIII methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* **82**, 143–153 (1995).
2. Pfeifer, G., Steigerwald, S., Hansen, R.S. Gartler, S.M. & Riggs, A.D. (1990) Polymerase-chain reaction aided genomic sequencing of an X chromosome linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl. Acad. Sci. USA* **87**, 8252–8256 (1990).
3. Iguchi-Ariga, S.M.M. & Schaffner, W. CpG methylation of the cAMP responsive enhancer/promoter TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* **3**, 612–619 (1989).
4. Bestor, T.H. & Coxon, A. The pros and cons of cytosine methylation *Curr. Biol.* **3**, 384–386 (1993).
5. Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl. Acad. Sci. USA* **92**, 7416–7419 (1995).
6. Herman, J. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA* **91**, 9700–9704 (1994).
7. Bestor, T.H. DNA methylation: evolution of a bacterial immune function into a regulator of genome structure and activity in higher eukaryotes. *Phil. Trans. Royal Soc. Lond.* **B 326**, 179–187 (1990).
8. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of mouse DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
9. Li, E., Beard, C., Forster, A.C., Bestor, T.H. & Jaenisch, R. DNA methylation, genomic imprinting, and mammalian development. *Cold Spring Harb. Symp. Quant. Biol.* LVIII, 297–305 (1993).
10. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226–232 (1975).
11. Riggs, A.D. X-inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9–11 (1975).
12. Bestor, T.H. & Verdine, G.L. DNA methyltransferases. *Curr. Opin. Cell Biol.* **6**, 380–389 (1994).
13. Bestor, T.H. The biochemical basis of allele-specific gene expression in genomic imprinting and X inactivation. in *Genomic Imprinting: Causes and consequences.* (eds Ohlsson, R., Hall, K. & Ritzen, M.) (Cambridge University Press, Cambridge, 1995).
14. Jaenisch, R., Schnieke, A. & Harbers, K. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proc. Natl. Acad. Sci. USA* **82**, 1451–1455 (1985).
15. Ashburner, M. *Drosophila: A Laboratory Handbook.* 303 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992).
16. Mummaneni, P., Bishop, P. & Turker, M. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. *J. Biol. Chem.* **268**, 552–558 (1993).
17. Macleod, D., Charlton, J., Mullins, J. & Bird, A. Sp1 binding sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* **8**, 2282–2292 (1994).
18. Brandeis, M. *et al.* Sp1 elements protect a CpG island from *de novo* methylation. *Nature* **371**, 435–438 (1994).
19. Craig, N.L. Unity in transposition reactions. *Science* **270**, 253–255 (1995).
20. Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**, 215–220 (1994).
21. Bestor, T.H. Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. *Nucl. Acids Res.* **15**, 3835–3843 (1987).
22. Chen, X. *et al.* Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* **92**, 5199–5203 (1995).
23. Jentsch, F. & Kemper, B. Endonuclease VII resolves Y-junctions in branched DNA *in vitro*. *EMBO J.* **5**, 181–190 (1986).
24. Lu, M., Guo, Q., Studier, W.F. & Kallenbach. N.F. Resolution of branched DNA substrates by T7 endonuclease I and its inhibition. *J. Biol. Chem.* **266**, 2531–2536 (1991).
25. Usdin, K. & Woodford, K.J. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucl. Acids Res.* **23**, 4202–4209 (1995).
26. Selker, E.U. Premeiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* **24**, 579–613 (1990).
27. Goyon, C., Nogueira, T.I.V. & Faugeron, G. Perpetuation of cytosine methylation in *Ascobolus immersus* implies a novel type of maintenance methylase. *J. Mol. Biol.* **240**, 42–51 (1994).
28. Matzke, M.A. & Matzke, A.J.M. Homology-dependent gene silencing in transgenic plants: what does it really tell us? *Trends Genet.* **11**, 1–3 (1995).
29. Matzke, A.J.M., Neuhuber, F., Park, Y.D., Ambros, P.F. & Matzke, M.A. Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes *Mol. Gen. Genet.* **244**, 219–229 (1994).
30. Flavell, R.B. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl. Acad. Sci. USA* **91**, 3490–3496 (1994).
31. Finnegan, J. & McElroy, D. Transgene inactivation: plants fight back! *Biotechnol.* **12**, 883–888. (1994)
32. Assaad, F.F., Tucker, K.L. & Signer, E.R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* **22**, 1067–1085 (1993).
33. Engels, W.R., Preston, C.R. & Johnson-Schlitz, D.M. Long-range cis preference in DNA homology search over the length of the Drosophila chromosome. *Science* **263**, 1623–1625 (1994).
34. Bender, J. & Fink, G. Epigenetic control of endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* **83**, 725–734 (1995)
35. Kricker, M.C., Drake, J.W. & Radman, M. Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proc. Natl. Acad. Sci. USA* **89**, 1075–1079 (1992).
36. Chesnokov, I.N. & Schmid, C.W. Specific Alu binding protein from human sperm chromatin prevents DNA methylation. *J. Biol. Chem.* **270**,18539–18542 (1995)
37. Rubin, C.M., VandeVoort, C.A., Tipfitz, R.L. & Schmid, C.W. Alu repeated DNAs are differentially methylated in primate germ cells. *Nucl. Acids Res.* **22**, 5121–5127 (1994).
38. Challita, P.-A. & Kohn, D.B. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*. *Proc. Natl. Acad. Sci. USA* **91**, 2567–2571 (1994).
39. Neumann, B., Kubicka, P. & Barlow, D.P. Characteristics of imprinted genes. *Nature Genet.* **9**, 451–452 (1994).
40. Hatada, I. & Mukai, T. Genomic imprinting of *p57KIP2*, a cyclin-dependent kinase inhibitor in mouse. *Nature Genet.* **11**, 204–206 (1995).
41. Barlow, D.P. Methylation and imprinting: from host defense to gene regulation? *Science* **260**, 309–310 (1993).
42. Michaud, E. *et al.* Differential expression of a new dominant agouti allele (A^{iapy}) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* **8**, 1463–1472 (1994).
43. Duhl, D., Vrieling, H., Miller, K., Wolff, G. & Barsh, G. Neomorphic agouti mutations in obese yellow mice. *Nature Genet.* **7**, 220–226 (1994).
44. Moulton, T. *et al.* Epigenetic lesions at the *H19* locus in Wilms' tumour patients. *Nature Genet.* **7**, 440–447 (1994).
45. Bulfing, K. *et al.* Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nature Genet.* **9**, 395–400 (1995).
46. Yang, A.S., Shen, J.-C., Zingg, J.-M., Mi, S. & Jones, P.A. Hhal and Hpall DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucl. Acids Res.* **23**, 1360–1367 (1995).
47. Kilmasauskas, S. & Roberts, R.J., M.Hhal binds tightly to substrates containing mismatches at the target base. *Nucl. Acids Res.* **23**, 1388–1395 (1995).
48. Jue, K., Bestor, T.H. & Trasler, J. Regulated synthesis and localization of DNA methyltransferase during spermatogenesis. *Biol. Reprod.* **53**, 561–569 (1995).
49. Bird, A.P. Gene number, noise reduction and biological complexity. *Trends Genet.* **11**, 94–100 (1995).
50. Cleaver, J. It was a very good year for DNA repair. *Cell* **76**, 1–4 (1994)
51. Taylor, J.H. Origins of replication and gene regulation. *Mol. Cell. Biochem.* **61**, 99–109 (1984).

TAB 4

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2567–2571, March 1994
Medical Sciences

Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*

(gene therapy/gene expression/bone marrow/long terminal repeat)

PIA-MARIA CHALLITA* AND DONALD B. KOHN*†‡

Departments of *Microbiology and †Pediatrics, University of Southern California School of Medicine, Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, Los Angeles, CA 90027

Communicated by Elizabeth F. Neufeld, November 18, 1993 (received for review July 1, 1993)

ABSTRACT **We describe studies of gene transfer and expression of the human glucocerebrosidase cDNA by a Moloney murine leukemia virus (MoMuLV)-based retroviral vector in a murine gene transfer/bone marrow transplant (BMT) model. Pluripotent hematopoietic stem cells (HSCs) were assayed as the colony-forming units, spleen (CFU-S) generated after serial transplantation. Transcriptional expression from the MoMuLV long-terminal repeat (LTR) was detected at a high level in the primary (1°) CFU-S and tissues of reconstituted BMT recipients. However, we observed transcriptional inactivity of the proviral MoMuLV-LTR in >90% of the secondary (2°) CFU-S and in 100% of the tertiary (3°) CFU-S examined. We have compared the methylation status of the provirus in the 1° CFU-S, which show strong vector expression, to that of the transcriptionally inactive provirus in the 2° and 3° CFU-S by Southern blot analysis using the methylation-sensitive restriction enzyme *Sma* I. The studies demonstrated a 3- to 4-fold increase in methylation of the *Sma* I site in the proviral LTR of 2° and 3° CFU-S compared to the transcriptionally active 1° CFU-S. These observations may have important implications for future clinical applications of retroviral-mediated gene transfer into HSCs, where persistent gene expression would be needed for an enduring therapeutic effect.**

Gene therapy via bone marrow cells is a promising technique for treatment of a wide variety of human diseases, including genetic disorders, cancer, and AIDS. Effective long-term bone marrow gene therapy requires the fulfillment of two main criteria. The exogenous gene should be introduced into a high percentage of long-lived pluripotent hematopoietic stem cells (HSCs). Subsequently, the introduced gene should be persistently expressed in the mature hematopoietic progeny cells of the stem cell, thereby maintaining the effects of gene therapy for the lifetime of the individual. Although Moloney murine leukemia virus (MoMuLV)-based retroviral vectors are currently the most efficient vehicles for gene transfer into a variety of cell types including HSCs (reviewed in ref. 1), the long-term *in vivo* expression from the viral promoter/enhancer elements has been unsatisfactory. Lack of gene expression from the 5′ MoMuLV long-terminal repeat (LTR) has been observed in several systems including primary fibroblasts (2) and hematopoietic cells (3, 4). Previous studies by our laboratory, using a retroviral vector in which a normal human glucocerebrosidase (GC) cDNA is controlled by the enhancer/promoter of the 5′ MoMuLV-LTR, demonstrated a high rate of lack of expression in cells derived from HSCs (5).

Methylation of cytosine residues has been shown to be associated with suppression of gene expression and, in certain circumstances, with the silencing of viral control elements (6). The MoMuLV-LTR is completely inactive in embryonic stem and embryonic carcinoma cell lines, and the inactivity is accompanied by *de novo* methylation of the proviral sequences (7, 8). Moreover, methylation has been detected in association with the MoMuLV-LTR transcriptional inactivity in fibroblasts *in vitro* (9) and *in vivo* (2).

In this study, we investigated long-term *in vivo* expression from the MoMuLV-LTR by transduction of murine bone marrow cells with a MoMuLV-based retroviral vector and serial bone marrow transplantation (BMT) into lethally irradiated recipient mice. We document a high rate of expression failure associated with methylation of the vector LTR in the secondary (2°) and tertiary (3°) colony-forming units, spleen (CFU-S).

MATERIALS AND METHODS

Retroviral Vector. The G2 retroviral vector and its corresponding high-titer amphotropic PA317 packaging cell clone have been described (5). G2 consists of the LTR from the N2 vector flanking the human GC cDNA. The packaging cell line clone used in the experiments was negative for helper virus production assayed by testing for transfer of the amphotropic *env* gene into 3T3 fibroblasts through PCR analysis (10).

Transduction of Murine Bone Marrow Cells. Donor bone marrow cells were harvested from male C57BL/6J mice (Charles River Breeding Laboratories), prestimulated in the presence of growth factors, and cocultivated over vector producing fibroblasts according to the methods described by Weinthal *et al.* (5). The growth factors used for the prestimulation were 200 units of murine interleukin 3 (IL-3) per ml (Biosource, Camarillo, CA), 100 units of human IL-6 per ml (Amgen), 200 units of human IL-1α per ml (Immunex) and 50 ng of mast cell growth factor per ml (or c-kit ligand; Immunex).

BMT and Sample Collection. Recipient female C57BL/6J mice (8–12 weeks old) were irradiated with two split doses of 600 and 450 cGy 24 hr apart. Transduced bone marrow cells were injected into the tail vein of the irradiated mice at 1×10^6 cells per mouse for isolation of CFU-S or $2–4 \times 10^6$ cells for long-term reconstitution. Twelve days after BMT, two to four mice transplanted with 1×10^6 bone marrow cells were sacrificed. Well-defined, individual primary (1°) CFU-S were isolated and divided evenly into two portions, one for DNA

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MoMuLV, Moloney murine leukemia virus; LTR, long terminal repeat; HSC, hematopoietic stem cell; GC, glucocerebrosidase; CFU-S, colony forming unit, spleen; BMT, bone marrow transplantation; 1°, primary; 2°, secondary; 3°, tertiary.

‡To whom reprint requests should be addressed at: Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA 90027.

Table 1. Expression of G2 in the mouse model of gene transfer/BMT

	Exp. 1		Exp. 2		Exp. 3		Total	
	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
1° CFU-S	9/9	5/5	12/12	6/6	8/8	8/8	29/29	19/19
1° tissues (1.5 months)								
Spleen					2/2	2/2	2/2	2/2
Thymus					2/2	2/2	2/2	2/2
Marrow					2/2	2/2	2/2	2/2
1° tissues (3 months)								
Spleen			1/1		2/2	2/2	2/2	2/2
Thymus					0/2	0/2	0/2	0/2
Marrow			1/1		2/2	2/2	3/3	2/2
2° CFU-S (1.5 months)	10/15	0/10*			3/28	0/3	13/43	0/13
2° CFU-S (3 months)			20/20	1/7	37/38	3/35	57/58	4/42
2° tissues (3 months)								
Spleen			2/2	1/2			2/2	1/2
Thymus			2/2	0/2			2/2	0/2
Marrow			2/2	1/2			2/2	1/2
3° CFU-S			7/29	0/7			7/29	0/7

*This set of 2° CFU-S was analyzed for RNA expression by reverse transcription/PCR.

and one for RNA analysis. Animals transplanted with 2–4 × 10^6 bone marrow cells were sacrificed after 1–3 months. Hematopoietic tissues were collected for nucleic acid analysis, and bone marrow cells were used directly to reconstitute a second generation of lethally irradiated female mice. Twelve days after the secondary BMT, the 2° CFU-S were isolated for DNA and RNA analysis. In one experiment, 3° BMT was performed from bone marrow of long-term-reconstituted 2° recipient animals in order to generate 3° CFU-S.

DNA and RNA Analysis. Genomic DNA was isolated by SDS/proteinase K and RNase digestion at 55°C for 3–4 hr. The digested tissues were extracted with phenol/chloroform; the DNA was precipitated in ethanol and resuspended in TE buffer. The presence of proviral GC sequences in the CFU-S and hematopoietic tissue samples was assayed by PCR using the human GC-specific oligonucleotide primers described by Weinthal *et al.* (5), followed by Southern blotting and hybridization with a ^{32}P-end-labeled internal oligonucleotide (8). Provirus DNA was also detected by Southern blot analysis after digestion of genomic DNA with the *Sst* II and *Xho* I restriction enzymes (BRL). These digestions release the 1.65-kb GC cDNA detected by hybridization with the 1.5-kb (*Sst* II/*Bam*HI) human GC cDNA probe. The probe was labeled with [^{32}P]dCTP by the random-priming method. Individual provirus integrants in the CFU-S and long-term hematopoietic tissues were detected by Southern blot analysis of genomic DNA digested with *Bam*HI, which cuts at one site in the provirus. Again, the Southern blot was hybridized with the 1.5-kb ^{32}P-labeled human GC cDNA probe.

RNA was isolated from the tissues by the acid guanidinium thiocyanate/phenol/chloroform method (11). RNA (15 μg) was electrophoresed on a 1.2% formaldehyde gel, denatured, neutralized, and transferred to a nylon membrane by capillary blotting. The filter was hybridized with the human GC cDNA probe. After a satisfactory exposure was obtained, the filter was stripped and rehybridized with the mouse β-actin DNA probe. For reverse transcription/PCR, 1 μg of RNA was reverse transcribed using the human GC-specific oligonucleotide primers, followed by PCR amplification of the cDNA as described above for the DNA samples.

Methylation Analysis. The methylation status of the proviral 5′ LTR in the CFU-S was determined by digestion of genomic DNA (15–25 μg) with *Bam*HI to reduce the size of the DNA fragments, followed by *Pvu* II digestion. The DNA was then precipitated with ethanol, redissolved in TE buffer, and divided into two equal portions, one of which was subjected to digestion with the methylation-sensitive enzyme *Sma* I. Completeness of the genomic DNA digestions was monitored by mixing a sample of the digestion mixture with either adenovirus type 2 DNA or λ DNA (BRL), which were subsequently run on a 1% gel. *Pvu* II and *Pvu* II/*Sma* I-digested DNA were electrophoresed and blotted to nylon membranes. The blots were probed with a ^{32}P-labeled fragment of the G2 vector from the *Spe* I site in the untranslated leader region to a *Pvu* II site near the 5′ end of the GC gene (see Fig. 3). Densitometric analyses were performed with the United States Biochemical SciScan 5000, measuring the relative densities of the 1.8-kb *Sma* I-resistant band and the 1.5-kb *Sma* I-sensitive band in each lane.

RESULTS

Expression of G2 *in Vivo* in Murine HSCs. Results of G2-mediated gene transfer and expression in the mouse

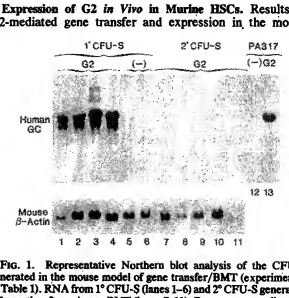


FIG. 1. Representative Northern blot analysis of the CFU-S generated in the mouse model of gene transfer/BMT (experiment 1 of Table 1). RNA from 1° CFU-S (lanes 1–6) and 2° CFU-S generated 1.5 months after primary BMT (lanes 7–11). Bone marrow cells used for BMT were transduced with G2 (lanes 1–4 and 7–11) or with the neomycin-containing control retroviral vector (lanes 5 and 6). RNAs from the fibroblast cell line PA317 (lane 12) and PA317 transduced with G2 (lane 13) were used as negative and positive controls, respectively, for GC mRNA. (*Upper*) Northern blot was probed with the human GC cDNA identifying the two proviral transcripts, the full-length 4.0 kb, and the spliced 3.5 kb. (*Lower*) Same blot was stripped and rehybridized with the mouse β-actin probe except for the control samples (lanes 12 and 13).

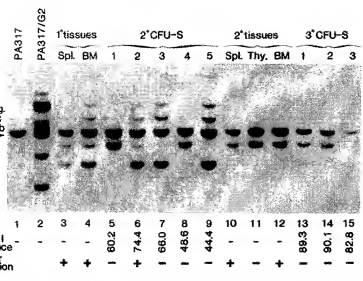


FIG. 2. Integration patterns of the G2 retroviral vector by Southern blot analysis. DNA was isolated from the tissues (experiment 2) and subjected to *Bam*HI digestion, which cuts at one site in the retroviral vector. 1° tissues refer to the spleen (lane 3) and bone marrow (lane 4) of one animal sacrificed 3 months after primary BMT. The marrow of this animal was used to reconstitute 2° recipients to generate the 2° CFU-S depicted in lanes 5–9. One reconstituted 2° recipient was sacrificed 3 months after 2° BMT. The spleen, thymus, and bone marrow (2° tissues) were analyzed (lanes 10–12). The marrow from this 2° BMT recipient was used in a 3° transplant to generate the 3° CFU-S (lanes 13–15). The extent of methylation measured for each CFU-S is indicated as the percentage of *Sma* I resistance. Vector expression was measured by Northern blot analysis and is indicated as present (+) or absent (−).

model of gene transfer/BMT are shown in Table 1. In the 1° CFU-S, a gene transfer extent of 100% was recorded in the three experiments performed. Furthermore, RNA expression was detected at a high level in all nineteen 1° CFU-S examined. An example of RNA expression from the Mo-MuLV-proviral LTR is shown in Fig. 1. The two proviral RNA transcripts in the 1° CFU-S are depicted in lanes 1–4. These data confirm previous findings from our laboratory, which demonstrated high-level expression by G2 in 1° CFU-S (5).

To confirm hematopoietic reconstitution with genetically marked cells, hematopoietic tissues were analyzed from mice 1–3 months after primary BMT. The intact human GC cDNA was detected by genomic DNA Southern blot analysis in all 10 spleen and marrow samples and in the thymus from two of four primary recipient animals. Vector RNA transcripts were detected, by Northern blot analysis, in all tissues in which the provirus DNA sequences were present (Table 1).

Analysis of the 2° CFU-S showed a high overall retroviral gene transfer frequency, with 70 of 101 (70%) containing the G2 provirus. However, RNA expression was rarely detected in these tissues. In a total of three experiments, RNA expression from the MoMuLV-LTR was detected in only 4 of 55 (7.3%) of the 2° CFU-S that contained the G2 provirus. The Northern blot in Fig. 1 shows an example of five 2° CFU-S that did not express proviral RNA transcripts (lanes 7–11). Experiment 2 was carried out further by allowing two other secondary BMT recipients to survive for 3 months after the transplantation, producing tissues stably engrafted by the serially passaged marrow (Table 1). The spleen, thymus, and marrow of the two reconstituted 2° recipients contained G2 provirus DNA by Southern blot. Vector RNAs were detected at a low level in the spleen and marrow, but not the thymus, of one of the two animals. No vector transcripts were detected in the other reconstituted 2° recipient. Tertiary BMT was also performed in experiment 2 with the bone marrow obtained from the reconstituted 2° recipients 3 months after secondary BMT. Seven of the twenty-nine 3° CFU-S analyzed contained G2 provirus DNA; these were all derived from the donor 2° animal, which had shown expression of vector RNA in its tissues. None of the seven 3° CFU-S containing proviral DNA had detectable vector RNA expression.

Gene Transfer into Pluripotent HSCs. Provirus integration patterns were followed in 1°, 2°, and 3° recipient animals to demonstrate that the 2° and 3° CFU-S, in which the LTR was inactive, are descended from true pluripotent HSCs (Fig. 2). Both the spleen and marrow of the 1° recipient, 3 months after BMT (Table 1, experiment 2), showed the same pattern, with five proviral integrants (Fig. 2, lanes 3 and 4). The five 2° CFU-S that were produced from marrow of that primary recipient showed segregation of the vector integrants seen in the primary tissues into two different patterns. The first pattern, represented in lanes 6, 7, and 9, consisted of four proviruses (although there may be slight contamination of the sample in lane 6 with DNA from a CFU-S with the integrant seen in lane 5). Interestingly, despite the same pattern of integrants among these three 2° CFU-S, there was strong vector expression seen in the one in lane 6, whereas the two foci seen in lanes 7 and 9 had no detectable vector transcripts. The second pattern of vector integrants seen in two other 2° CFU-S (lanes 5 and 8) had one provirus. The latter single integrant of ≈4 kb was also detected in the spleen, thymus, and marrow of another reconstituted secondary animal 3 months after secondary BMT (lanes 10–12). Moreover, this integrant was transferred to three CFU-S of a 3° recipient animal. Thus, the same transduced stem cell that generated the 2° CFU-S was able to reconstitute all three hematopoietic tissues (spleen, thymus, and marrow) of a secondary animal sacrificed 3 months after BMT and also to generate 3° CFU-S in the 3° transplanted animal. The provirus present in the 2° and 3° CFU-S was transcriptionally inactive (Table 1).

Methylation Analysis of the CFU-S. Genomic DNA from transcriptionally active 1° CFU-S and from inactive 2° and 3° CFU-S were compared for *Sma* I resistance according to the protocol described in *Materials and Methods* and illustrated in Fig. 3. *Sma* I is a methylation-sensitive enzyme that can cleave DNA at the CCCGGG site only if the CpG sequence is not methylated; therefore, *Sma* I resistance is used as a measurement of DNA methylation. Two examples of the resulting Southern blots are shown in Fig. 4. In the packaging cell line PA317, the G2 provirus gives a 1.8-kb *Pvu* II band (Fig. 4, lanes 1), which is reduced to a 1.5-kb band upon *Sma* I digestion (lanes 2). Complete digestion by *Sma* I shows the absence of methylation of the 5′ LTR in the fibroblast cell

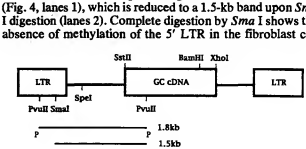


FIG. 3. Map of the G2 retroviral vector showing restriction sites and probe used for DNA methylation analysis.

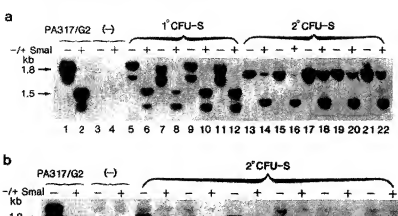


FIG. 4. Southern blot analysis of methylation of the G2 provirus in the 1° CFU-S (*a*) and 2° CFU-S (*a* and *b*). DNA was subjected to *Bam*HI and *Pvu* II with or without *Sma* I digestions. Lanes 3 and 4, labeled (−), represent DNA from normal control C57BL/6J mice. A second smaller band is seen in the PA317 cells and in the 1° CFU-S analyzed here due to the presence of a provirus with a short deletion in the 5′ untranslated region among the six copies in the packaging cell clone.

line. All *Sma* I digestions of DNA from 1° CFU-S show reduction in size of the vector-specific band to 1.5 kb, corresponding to lack of methylation of the *Sma* I site of the provirus 5′ LTR in 1° CFU-S. In contrast, the provirus in all 2° CFU-S shows some extent of *Sma* I resistance reflecting various degrees of methylation of the 5′ MoMuLV-LTR in these tissues.

To quantitate the degree of methylation at the *Sma* I site in the CFU-S, densitometric analyses were performed on the Southern blots, comparing the relative intensities of the 1.8- and 1.5-kb bands. Fig. 5 displays the percentage *Sma* I resistance of the G2 provirus in 3T3 fibroblasts, 1° CFU-S, and 2° and 3° CFU-S. In experiment 1 (Fig. 5*A*), the 1° CFU-S showed a mean of 8.9% *Sma* I resistance. In contrast, the average percentage *Sma* I resistance recorded in the 2° CFU-S generated 1.5 months after initial gene transfer/BMT was 24.6%, ranging from 9.0% to 47.6% over 10 samples analyzed. Thus, a 3-fold increase in percentage *Sma* I resistance was recorded between the 1° and 2° CFU-S of this experiment. The difference in methylation is statistically significant, with $P < 0.005$ by Student's *t* test analysis. Fig. 5*B* represents the comparison between 1° and 2° CFU-S derived 3 months after initial gene transfer/BMT in experiment 2. The percentage *Sma* I resistance was 14.9% in the 1° CFU-S and increased 3.8-fold to 56.8% in the 2° CFU-S. In the same experiment, the 3° CFU-S generated 3 months after secondary BMT showed 74.7% *Sma* I resistance, representing a 3.9-fold increase over the 19.3% *Sma* I resistance seen in other 1° CFU-S from the same experiment (Fig. 5*C*). Methylation of the provirus in both 2° and 3° CFU-S of this experiment was significantly greater than in 1° CFU-S ($P < 0.001$ by Student's *t* test). We conclude that the transcriptional inactivity observed in the 2° and 3° CFU-S is associated with methylation of the 5′ MoMuLV-provirus LTR at the *Sma* I site, 30 bp downstream of the transcription start site. Of note, the methylation status of the three 2° CFU-S with the common pattern of four vector integrants (Fig. 2, lanes 6, 7, and 9) was similar, ranging from 44% to 74%, despite the discordance for expression.

DISCUSSION

We have studied gene transfer of a MoMuLV-based retroviral vector into murine hematopoietic stem cells and expression from the MoMuLV-LTR promoter/enhancer elements in the progeny of the transduced cells. The expression from the proviral LTR was measured in the CFU-S derived from primary BMT performed after gene transfer, in the hematopoietic tissues of long-term reconstituted animals, and in the CFU-S generated after serial transplantation.

Our results demonstrate that the MoMuLV-LTR is a very efficient expression unit in 1° CFU-S. We have also detected expression from the MoMuLV-LTR in the hematopoietic tissues of transplant recipients 3 months after primary BMT.

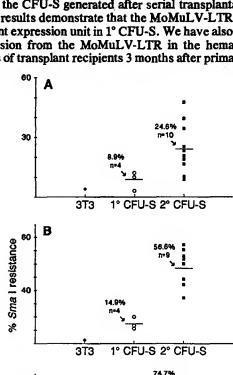


FIG. 5. Densitometric analysis to quantitate the extent of methylation of the MoMuLV-LTR in the 1°, 2°, and 3° CFU-S. The densities of the 1.8- and 1.5-kb bands after *Sma* I digestion were measured and percentage *Sma* I resistance was calculated. Each number represents the mean of three different readings. (*A*) Percentage *Sma* I resistance of samples from experiment 1. (*B* and *C*) Results from experiment 2 of the 2° CFU-S generated 3 months after secondary BMT and 3° CFU-S generated 3 months after 3° BMT, respectively. In each graph, the 1° CFU-S and either the 2° or the 3° CFU-S were analyzed from the same blot.

These results are comparable with previously published observations (4, 5, 12, 13). However, 1° CFU-S are mainly derived from committed progenitor cells restricted to the myeloid lineage (14, 15). Even the hematopoietic cells in the organs 3 months after BMT may be derived from progenitor cells capable of short-term but not long-term engraftment (16). By following the proviral integration pattern after serial transplantation, we were able to show that the cells capable of forming 2° CFU-S have the characteristics of long-lived, pluripotent HSCs. In this stringent assay of HSCs, we have observed that the MoMuLV-LTR is frequently inactive in the resultant progeny cells derived from the HSCs. We cannot, however, determine whether the absence of expression in the 2° CFU-S reflects the loss of expression by the LTR with time or whether the LTR is never active if inserted into the subclass of long-lived HSCs capable of producing 2° and 3° CFU-S.

The failure of transcription from the MoMuLV-LTR in hematopoietic tissues is in accord with prior observations (3–5, 17). However, the previous studies have mostly analyzed expression in reconstituted hematopoietic tissues of primary recipients. Moore *et al.* (17) have studied expression of human adenosine deaminase (ADA) by the MoMuLV-LTR vector expression through 2° CFU-S. Despite strong ADA expression in the primary recipients, they detected expression in only eighteen of seventy-two 2° CFU-S, although the percentage of these colonies that contained proviral DNA was not determined.

Methylation is associated with transcriptional inactivation of many genes and has specifically been seen in association with inactivity of the transduced MoMuLV-LTR in embryonic stem and embryonic carcinoma cell lines (7). Therefore, we examined the methylation status of the inactive 5′ MoMuLV-LTR in the 2° and 3° CFU-S. The presence of a high copy number of endogenous murine retroviral sequences creates a high background in the analysis of methylation across the provirus 5′ LTR. To overcome this problem, we used the restriction enzymes *Sma* I and *Pvu* II, which generate a specific provirus band detected by Southern blot analysis. Unfortunately, this assay restricted our analysis of methylation to one CpG dinucleotide contained in the *Sma* I site, 30 bp downstream from the transcription start site in the 5′ MoMuLV-proviral LTR.

Our analysis has indicated striking differences in the methylation patterns of this sequence. The *Sma* I site described is extensively methylated in the 2° CFU-S, which do not show MoMuLV proviral transcription, but is not methylated in the 1° CFU-S, which do express vector transcripts. We have observed that this same *Sma* I site is completely methylated in vector-transduced embryonic stem cell lines, which also do not show expression from the MoMuLV-LTR (data not shown). Moreover, studies done by Singer-Sam *et al.* (18) on the phosphoglycerate kinase promoter present on the inactive X chromosome have shown that methylation of a similar site, a *Hpa* II site (CCGG) at position +20, correlates with lack of transcription from the promoter. Nevertheless, the observed association between proviral methylation and expression inactivity does not show whether methylation plays a causal role in suppressing expression or is merely a secondary event after failure of expression has occurred.

Interestingly, the association between methylation and expression inactivity was not complete; at least one 2° CFU-S with extensive methylation at the *Sma* I site had a high level of vector transcripts. Other 2° CFU-S with the same proviral integrants failed to express the vector. This set of samples would indicate that neither integration site, methylation at the *Sma* I site, nor differentiation status of the transduced stem cell act to absolutely govern expression. The observed discordance suggests that commitment to expression or inactivity may stochastically be made after the HSC has differentiated to produce multiple pre-CFU-S.

The studies presented in this paper suggest that the wild-type MoMuLV-LTR may not be the ideal transcriptional unit for expression in pluripotent HSCs and their progeny cells. Successful approaches to overcoming this problem have been to use either modified LTRs with enhancer substitutions (19) or internal promoters derived from housekeeping genes (such as phosphoglycerate kinase or β-actin) instead of the wild-type MoMuLV-LTR (20, 21). Characterization of a transcriptionally active retroviral vector in HSCs may provide a better understanding of the regulation of gene expression occurring in these cells.

We would like to thank Dr. Robertson Parkman, Dr. Gay Crooks, and Dr. Punam Malik for carefully reading the manuscript; Dr. Wanda Krall for her helpful advice; and Dr. Ellen Bolotin for assistance in statistical analysis. The study was supported by National Institutes of Health Grant DK-42694 and March of Dimes Grant 1-FY92-0551. D.B.K. is a recipient of an American Cancer Society Junior Faculty Research Award (JFRA-320).

1. Mulligan, R. C. (1993) *Science* **260**, 926–932.
2. Palmer, T. D., Rosman, G. J., Osborne, W. R. & Miller, A. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1330–1334.
3. Williams, D. A., Orkin, S. H. & Mulligan, R. C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2566–2570.
4. Kaleko, M., Garcia, J. V., Osborne, W. R. & Miller, A. D. (1990) *Blood* **75**, 1733–1741.
5. Weinthal, J., Nolta, J. A., Yu, X. J., Lilley, J., Uribe, L. & Kohn, D. B. (1991) *Bone Marrow Transplant.* **8**, 403–412.
6. Cedar, H. (1988) *Cell* **53**, 3–4.
7. Habers, K., Schnieke, A., Stuhlman, H., Jahne, D. & Jaenisch, R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7609–7613.
8. Jaehner, D. & Jaenisch, R. (1985) *Nature (London)* **315**, 594–597.
9. Hoeben, R. C., Migchielsen, A. A. J., van der Jagt, R. C. M., van Ormondt, H. & van der Eb, A. J. (1991) *J. Virol.* **65**, 904–912.
10. Morgan, R. A., Cornetta, K. & Anderson, W. F. (1990) *Hum. Gene Ther.* **1**, 135–149.
11. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
12. Karlsson, S. (1991) *Blood* **78**, 2481–2492.
13. Krall, W. J., Challita, P. M., Perlmutter, L. S., Skelton, D. C. & Kohn, D. B. (1994) *Blood*, in press.
14. Ploemacher, R. E. & Brons, R. H. (1989) *Exp. Hematol.* **17**, 263–266.
15. Jones, R. J., Wagner, J. E., Celano, P., Zicha, M. S. & Sharkis, S. J. (1990) *Nature (London)* **347**, 188–189.
16. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. (1986) *Cell* **45**, 917–927.
17. Moore, K. A., Fletcher, F. A., Villalon, D. K., Utter, A. E. & Belmont, J. W. (1990) *Blood* **75**, 2085–2092.
18. Singer-Sam, J., Goldstein, L., Dai, A., Gartler, S. M. & Riggs, A. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1413–1417.
19. van Beusechem, V. W., Kukler, A., Einerhand, M. P. W., Bakx, T. A., van der Eb, A. J., van Bekkum, D. W. & Valerio, D. (1990) *J. Exp. Med.* **172**, 729–736.
20. Lim, B., Apperley, J. F., Orkin, S. H. & Williams, D. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8892–8896.
21. Guild, B. C., Finer, M. H., Housman, D. E. & Mulligan, R. C. (1988) *J. Virol.* **62**, 3795–3801.

TAB 5

MOLECULAR AND CELLULAR BIOLOGY, Oct. 2000, p. 7419–7426 Vol. 20, No. 20
0270-7306/00/$04.00+0
Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Retroviral Expression in Embryonic Stem Cells and Hematopoietic Stem Cells

SARA R. CHERRY,[1,2] D. BINISZKIEWICZ,[1] L. VAN PARIJS,[3] D. BALTIMORE,[3] AND R. JAENISCH[1,2]*

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142[1]; Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[2]; and California Institute of Technology, Pasadena, California 91125[3]

Received 5 May 2000/Accepted 19 June 2000

Achieving long-term retroviral expression in primary cells has been problematic. De novo DNA methylation of infecting proviruses has been proposed as a major cause of this transcriptional repression. Here we report the development of a mouse stem cell virus (MSCV) long terminal repeat-based retroviral vector that is expressed in both embryonic stem (ES) cells and hematopoietic stem (HS) cells. Infected HS cells and their differentiated descendants maintained long-term and stable retroviral expression after serial adoptive transfers. In addition, retrovirally infected ES cells showed detectable expression level of the green fluorescent protein (GFP). Moreover, GFP expression of integrated proviruses was maintained after in vitro differentiation of infected ES cells. Long-term passage of infected ES cells resulted in methylation-mediated silencing, while short-term expression was methylation independent. Tissues of transgenic animals, which we derived from ES cells carrying the MSCV-based provirus, did not express GFP. However, treatment with the demethylating agent 5-azadeoxycytidine reactivated the silent provirus, demonstrating that DNA methylation is involved in the maintenance of retroviral repression. Our results indicate that retroviral expression in ES cells is repressed by methylation-dependent as well as methylation-independent mechanisms.

Retroviral vectors are appealing vehicles for gene transfer. However, long-term expression mediated by integrated proviruses in primary cells has been difficult to achieve. Retroviral regulatory elements are repressed in numerous cell types, including embryonic stem (ES) cells and hematopoietic stem (HS) cells (1, 3). For example, vectors that are functional in mature hematopoietic cells are often not expressed in blood cells of animals transplanted with the infected stem cells (18, 19, 31). In particular, the lack of significant provirus transcription in ES cells and their differentiated descendants has hampered the use of retroviral vectors in transgenic experiments (5, 12, 32). Interestingly, this block in provirus expression is maintained upon differentiation of infected cells despite the fact that primary infection of cells after differentiation results in efficient expression (6, 7, 26).

Transcriptional repression is thought to be mediated by both *cis*-acting de novo methylation of the integrated proviruses and cell-type-specific *trans*-acting transcriptional repressors (5, 9, 23). The effect of *trans*-acting factors on retroviral expression through binding of specific sequences within the promotors of retroviruses has been examined in many studies (29, 30, 35). In fact, the mouse stem cell virus (MSCV) long terminal repeat (LTR) was generated by the modification of the sequences within the LTR to increase the affinity for positive factors and decrease the affinity for negative regulators (20).

In contrast, the role of methylation in silencing has been less clear. DNA methylation is thought to be a general mechanism used by cells to silence foreign DNA and may be involved in the cell defense against transposable elements (39). DNA methylation has also been associated with the repression of gene expression and the silencing of viral control elements (2, 14, 38). Exogenously introduced retroviruses silenced in vitro and in vivo can be reactivated by treatments that result in genomewide demethylation. In addition, transcriptionally silent endogenous retroviral elements are reactivated upon loss of genomic methylation in *Dnmt1* knockout mice (38). Therefore, DNA methylation is thought to causally repress expression of retroviral promoters in a variety of cell types.

ES cells provide a good model to study the role of DNA methylation in retroviral silencing. First, it was demonstrated that ES cells have high de novo methylation activity, which leads to effective methylation of integrated retroviral vectors, while little or no de novo methylation activity was detected in differentiated cells (21). In addition, ES cells were genetically modified to alter the endogenous level of DNA methylation by the targeted disruption of the maintenance methyltransferase gene *Dnmt1*. ES cells homozygous for this mutation proliferate normally with their genomic DNA highly demethylated, while differentiated cells and mice die due to the loss of genomic methylation (21, 22). Therefore, these modified ES cells are useful to study the effect of DNA methylation on retroviral gene expression. In addition, ES cells can be induced to differentiate in vitro or in vivo, allowing the study of DNA methylation and its effect on long-term expression.

Both Moloney virus-based and MSCV-based retroviral vectors have been used for gene transduction in a variety of cells. The MSCV vector is different from the typical Moloney virus vector in that the mutations in the LTR have allowed expression in a larger host range (8, 20). To this end, we modified MSCV to express the green fluorescent protein (GFP) as a sensitive reporter for gene expression (37). Using this vector, we demonstrated efficient expression in both ES and HS cells. We also demonstrated that silencing of retroviruses involves two mechanisms: (i) *trans*-acting factors that affect the initial expression of Moloney virus-based vectors but not MSCV-based vectors and (ii) long-term DNA methylation-dependent silencing that directly restricts expression of MSCV in ES cells and during embryogenesis. Silencing of the MSCV vector in wild-type ES cells and in in vivo differentiated ES cells was

* Corresponding author. Mailing address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142. Phone: (617) 258-5186. Fax: (617) 258-6505. E-mail: jaenisch@wi.mit.edu.

Downloaded from mcb.asm.org by on September 14, 2009

reversed by 5-azadeoxycytidine (5-azadC) treatments that demethylated the retroviral sequences, demonstrating that DNA methylation directly controls the maintenance of retroviral repression.

MATERIALS AND METHODS

Tissue culture. ES cells were cultured as described previously (21). To generate ES cell clones for injection into blastocysts, the ES cells were maintained on irradiated mouse embryonic fibroblasts (MEFs) with 500 U of leukemia inhibitory factor (LIF) per ml (22). For other experiments, the ES cells were cultured without MEFs in 1,000 U of LIF per ml. 293 cells were maintained in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum (FBS), penicillin, streptomycin, and glutamine. Abelson virus-transformed B cells were maintained in RPMI 1640 supplemented with 10% defined FBS (HyClone), penicillin, streptomycin, glutamine, and 50 μM β-mercaptoethanol. ES cells with retroviral integrants were in vitro differentiated as follows: the cells were passaged without LIF in the absence of MEFs on bacterial plastic petri dishes for 4 days, trypsinized, and cultivated with or without retinoic acid for 2 weeks (25).

Plasmids. The retroviral vectors MfgGFP, pMXGFP, and MSCVIresGFP have been described elsewhere (27, 33, 37). The MSCVIresGFP vector was modified by introducing either the Cre recombinase or the human Bcl-2 gene upstream of the internal ribosome entry site (IRES)-GFP cassette as described elsewhere (11, 37). The replication-incompetent helper plasmid pCL-eco was used (24).

Retroviral infections. To generate retroviral supernatants, 293 cells were transiently transfected by calcium phosphate-mediated coprecipitation with 5 μg of the replication-incompetent helper vector pCL-eco and 10 μg of the reporter retroviral vector as stated elsewhere (28). The cells were fed at 24 h postinfection, and the retroviral supernatant was used at 48 h. The cells continued to produce high-titer retroviruses for another 2 days, and that supernatant was used if needed for additional experiments. The supernatant was collected, brought to 4 μg of Polybrene per ml–10 mM HEPES, and filtered (0.45-μm-pore-size filter) for use.

ES cells for infection were washed and trypsinized. They were then plated at 10^6 cells per well of a six-well dish and centrifuged. The ES cell medium was removed, and retroviral supernatant was added at 1 ml/10^6 cells. Next, the plate was centrifuged for 45 min at 2,500 rpm at room temperature. The retroviral supernatants were removed; the cells were resuspended in ES cell medium and plated onto gelatinized dishes. ES cells used to generate mice were plated onto irradiated MEFs.

Bone marrow was infected as follows (36). C57BL/6 mice were obtained from the Jackson Laboratory (Bar Harbor, Maine). Bone marrow cells were harvested from the tibias and femurs of C57BL mice 5 days after they received an intraperitoneal injection of 5 mg of 5-fluorouracil (Sigma) in Dulbecco's phosphate-buffered saline (Gibco/BRL). These cells were then cultured for 4 days at 2×10^6 cells/ml with recombinant mouse interleukin-3 (rmIL-3; 20 ng/ml), rmIL-6 (50 ng/ml), and (50 ng/ml; recombinant mouse stem cell factor; R&D Systems) in Dulbecco's modified Eagle medium containing 10% FBS. After 48 and 72 h, the bone marrow cells were spin infected with the retroviral supernatant generated as described above. Then the retroviral supernatant was removed and replaced with growth medium containing cytokines.

FACS (fluorescence-activated cell sorting) analysis and sorting. Adherent cell lines were trypsinized, washed, and resuspended in complete medium to achieve a single-cell suspension at the time points indicated. Nonadherent cells were used directly for analysis. Organs were disrupted manually and passed through a 70-μm mesh to generate a single-cell suspension. The cells were analyzed for viability using scatter properties and the exclusion of propidium iodide. The level of GFP expression was monitored by fluorescence without compensation to detect cells with low levels of GFP expression. The ES cells were sorted into ES cell medium and plated immediately onto either gelatinized plates or MEFs for blastocyst injections. The survival of ES cells after sorting was approximately 50%, as measured by the number of colonies generated divided by the expected number of colonies.

5-azadC treatments. ES cells were treated with 0.15 μM 5-azadC (Sigma) at days 1 and 3 postplating. The cells were fed, allowed to recover, and then assayed 4 to 8 days later. The red blood cells in whole blood were lysed (5), and the remaining cells were stained with fluorescently labeled anti-H2-b, anti-H2-d, anti-B220, anti-TCRα (Pharmingen) at 1:200 as indicated. At day 0, splenocytes were treated with either anti-CD3 or anti-CD40 (Pharmingen); 0.15 μM 5-azadC was added at day 1, and the anti-CD3-treated cells were assayed at day 4. 5-AzadC was added again to the B-cell cultures with fresh anti-CD40 at day 4, and the cells were assayed at day 6.

Staurosporine-mediated cell death. ES cells were infected with the stated retrovirus and treated with staurosporine at day 4 postinfection for 24 h with the indicated concentration of drug. The percentage of viable, GFP-positive cells was determined by flow cytometry (6). Data are presented as a percentage of GFP-positive cells before treatment. Results from one representative experiment of three performed are shown.

LacZ staining. ES cells were infected with the stated retrovirus and sorted for GFP expression at day 3 postinfection. The ES cells were plated and cultured for an additional 5 days and stained for LacZ expression as described elsewhere (41).

Adoptive transfers. Recipient mice (10) received a total of 1,200 rads of whole-body radiation in two doses (800 and 400 rads) 3 h apart and were then injected with 2×10^6 to 5×10^6 infected bone marrow cells. Irradiated mice were maintained on trimethaprim-sulfamethoxazole in sterile cages for 4 to 6 weeks to prevent opportunistic infections (34). Serial passages were performed by harvesting bone marrow from mice 6 to 8 weeks postreconstitution and transferring 2×10^6 to 5×10^6 cells into irradiated recipients. Mice were analyzed 8 to 12 weeks posttransfer to allow reconstitution of the T-cell compartment. These experiments were repeated multiple times with similar results.

Southern blot analysis. The genomic DNA was isolated as described elsewhere (19). Ten micrograms of DNA was digested with the stated restriction enzyme overnight. The products were resolved on an agarose gel, transferred to a nylon membrane, and detected using a probe that spans the entire GFP coding sequence.

RESULTS

High-efficiency retroviral expression in ES cells. Retroviral vectors based on the MSCV LTR were constructed with a multiple cloning site followed by an IRES driving expression of the gene for GFP as schematically diagrammed in Fig. 1A (MiG) (37). We generated high-titer retroviruses by transient transfection and infected ES cells with an adapted spin infection protocol. Using this protocol, we reproducibly achieved high-efficiency (>50%) infection of ES cells as measured by flow cytometry; uninfected control cells were negative for GFP expression (Fig. 1B). The intracellular concentration of GFP is directly proportional to the fluorescence intensity measured by flow cytometry.

Next, we compared expression of the MSCV-based retrovirus and Moloney virus-based retroviral vectors in ES cells. GFP expression was detectable with the MSCV LTR-containing MiG vector but not with the two Moloney virus-based viruses pMX (27) and Mfg (33) (Fig. 1B). This was not due to inefficient genomic integration of the provirus or to a lower titer. Southern blot analysis of genomic DNA demonstrated that all three proviruses were integrated in the ES cells (Fig. 1C). Also, when parallel B-cell cultures were infected with the retroviral supernatant used to infect ES cells, all of the retroviral vectors were expressed in B cells at comparable efficiencies (Fig. 1B).

GFP expression driven by the MSCV LTR in ES cells was substantially lower than in other differentiated cell lines tested (Fig. 1B and data not shown) (20). To determine whether this low level of expression was sufficient to drive functional expression of other gene products, we cloned the Cre recombinase upstream of the IRES-GFP cassette to generate MSCVCreiresGFP. We tested for Cre activity by infecting ES cells that contain a translational stop sequence flanked by *loxP* sites located between the Rosa 26 promoter and a *lacZ* reporter schematically diagrammed in Fig. 2A (34). If Cre is expressed at functional levels in these ES cells, the protein will catalyze recombination of the *loxP* sites, leading to loss of the stop sequences and the expression of LacZ. Indeed, we found that >99% of GFP-positive ES cells that were infected with the Cre-expressing retrovirus were also LacZ positive (Fig. 2B). Uninfected cells were both GFP negative and LacZ negative (data not shown). This indicates that virus-mediated gene transfer resulted in functional Cre expression.

Because Cre activity is required only transiently for LacZ expression, we tested a second gene product that must be stably expressed throughout the experiment. It has been demonstrated that Bcl-2 expression protects many cell types against staurosporine-mediated apoptosis (10). Therefore, we examined whether Bcl-2 could protect ES cells from apoptosis when expressed from the MSCV LTR. We cloned human Bcl-2 up-

Downloaded from mcb.asm.org by on September 14, 2009

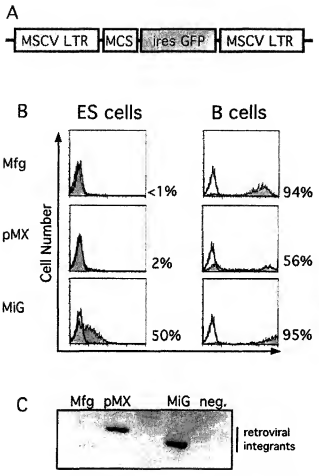


FIG. 1. Efficient retroviral infection of ES cells. (A) Schematic diagram of MiG vector containing the MSCV LTR followed by a multiple cloning site (MCS) and an IRES-GFP cassette. (B) MSCV-based (MiG) but not Moloney virus-based (Mfg and pMX) retroviruses express in ES cells. B cells or ES cells were infected by the indicated retroviruses and assayed by flow cytometry 2 days postinfection. Uninfected cells (unshaded) and infected cells (shaded area) were electronically gated for live cells and subsequently analyzed for GFP fluorescence and for cell number. Percentages of GFP-positive cells are indicated. (C) Comparable levels of integration of different retroviruses into ES cells, determined by Southern blot analysis of genomic DNA purified from infected and uninfected ES cells 2 days postinfection, digested with *Kpn*I, a restriction site present within the LTRs, and probed with the GFP coding sequence.

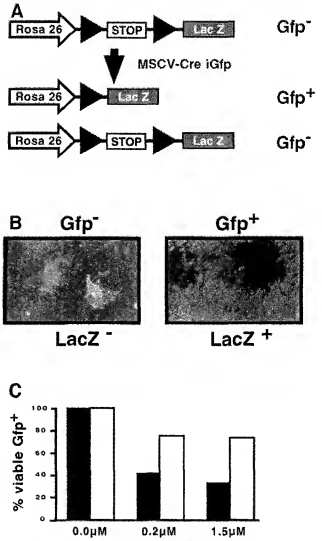


FIG. 2. Expression from the MSCV LTR is sufficient to drive functional gene expression. (A) Schematic diagram of the Rosa 26 locus in Cre reporter ES cells. Before Cre-mediated recombination, LacZ expression is prevented by the presence of a stop fragment. Retroviral infection with a Cre-expressing retrovirus with a GFP reporter results in two populations of cells. Cells that are GFP$^+$ become LacZ$^+$ due to efficient Cre-mediated recombination of the stop fragment. In contrast, cells that are GFP$^-$ were not infected and thus remained LacZ$^-$. (B) ES cells were infected with the MSCVCreiresGFP retrovirus and sorted for either Gfp$^-$ or Gfp$^+$ as indicated. The cells were subsequently cultured and stained for LacZ expression. Gfp$^-$ cells are white (and therefore LacZ$^-$) while Gfp$^+$ cells are blue (and therefore LacZ$^+$). More than 99% of the Gfp$^+$ cells were LacZ$^+$ in multiple experiments. (C) ES cells were infected with either MSCViresGFP (■) or MSCVBcl-2iresGFP (□) and treated with the indicated amounts of staurosporine. The percentage of viable, Gfp$^+$ (infected) cells was determined by flow cytometry. The results are shown as a percentage of Gfp$^+$ cells before treatment. The results are from one representative experiment of three performed.

stream of the IRES-GFP cassette to generate MSCV Bcl-2iresGFP. Wild-type ES cells were infected with either the Bcl-2-expressing retrovirus or the control virus lacking Bcl-2. Increasing concentrations of staurosporine were added to the cultures, and flow cytometry was used to assay for both viability and GFP expression. GFP-positive cells infected with the Bcl-2-containing virus were significantly protected from staurosporine-mediated cell death compared to the GFP-negative cells or GFP-positive cells infected with the control retrovirus (Fig. 2C). Therefore, the level of expression from the MSCV LTR is sufficient for stable functional gene expression in ES cells.

Short-term transcriptional silencing in ES cells is methylation independent. It long has been hypothesized that retroviruses are transcriptionally silenced in embryonic cells by DNA methylation (12, 14, 21). Therefore, it was possible that DNA methylation of the MSCV LTR was responsible for the decreased level of expression in ES cells compared to other cell types (Fig. 1B). In addition, we sought to test whether DNA methylation of the Moloney virus-based vectors in the wild-type ES cells was the mechanism by which the Moloney virus-based LTRs were silenced (9, 13). To this end, we infected ES cells deficient for the maintenance DNA methyltransferase gene, *Dnmt1*, the loss of which results in genomewide hypomethylation (21, 22). *Dnmt1*$^{-/-}$ ES cells are demethylated, and proviral sequences remain unmethylated. The Moloney virus-based retroviruses such as pMX remained silent even when

Downloaded from mcb.asm.org by on September 14, 2009

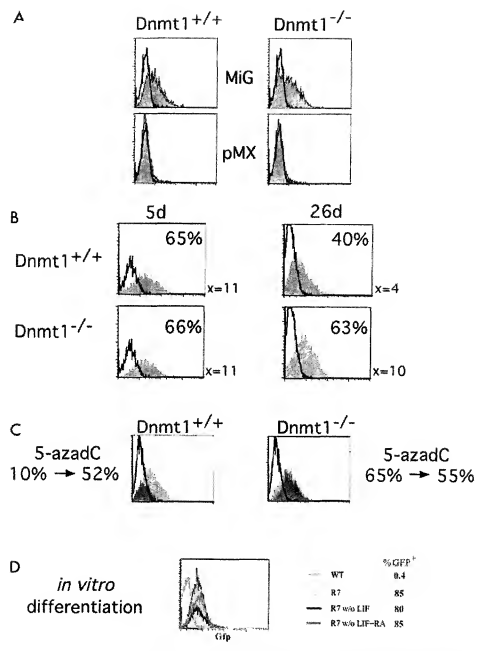


FIG. 3. Long-term expression of retroviruses is repressed by methylation. (A) MSCV-based (MiG) but not Moloney virus-based (pMX) retroviruses express in ES cells independent of the methylation status of the cells. *Dnmt1*$^{+/+}$ or *Dnmt1*$^{-/-}$ ES cells were not infected (unshaded) or infected by the indicated retroviruses (shaded) and assayed by flow cytometry 2 days postinfection as for Fig. 1B. (B) Long-term expression of GFP in ES cells is decreased in *Dnmt1*$^{+/+}$ cells but not *Dnmt1*$^{-/-}$ cells. The ES cells were infected with MiG, passaged for 5 or 26 days postinfection, and assayed by flow cytometry as above. The mean fluorescent intensity for the population and the percentage of GFP-positive cells are indicated. (C) Treatment with 5-azadC rescues the expression of retroviruses silenced by long-term passage. *Dnmt1*$^{+/+}$ or *Dnmt1*$^{-/-}$ ES cells were infected by MiG and passaged for >40 days. The cells were divided, and half were treated with 5-azadC. Then uninfected ES cells (unshaded), MiG-infected untreated ES cells (dark shading), and MiG-infected 5-azadC-treated ES cells (light shading) were assayed by FACS analysis. Numbers below the FACS plots are percentages of GFP-positive cells before and after 5-azadC treatment. (D) In vitro differentiation of ES cells does not affect retroviral expression. A clonal ES cell line (R7) infected with MiG or an uninfected wild-type (WT) ES cell control was in vitro differentiated by passage without feeders and LIF, with or without retinoic acid (RA) as indicated. The cells were assayed by flow cytometry, and the percentage of GFP-positive cells is indicated.

introduced into *Dnmt1*$^{-/-}$ ES cells, whereas MSCV expressed similar levels of GFP in both *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ ES cells (Fig. 3A). Therefore, the initial block in transcription directed by Moloney virus LTRs in ES cells is independent of DNA methylation and is presumably due to the binding of *trans*-acting factors. In addition, the mean fluorescence intensities of GFP were comparable between the *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ ES cells, indicating that the basal level of expression of the MSCV LTR is independent of DNA methylation.

DNA methylation constrains long-term retroviral expression. MiG-infected GFP-expressing ES cells were continually passaged to test the effect of DNA methylation on long-term expression.

Downloaded from mcb.asm.org by on September 14, 2009

Though GFP expression was high in both *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ ES cells at 5 days postinfection, a substantial fraction of the infected wild-type ES were GFP negative at 26 days postinfection. This was apparent by both a loss in the percentage of GFP-positive cells as well as a decrease in the mean fluorescence intensity of the bulk population of wild-type ES cells and was observed in both bulk cultures and individual cloned lines containing single integrants (Fig. 3B and data not shown). The fraction of GFP-positive cells continues to decrease with additional passages, as shown in Fig. 3C. These results suggest that long-term expression was suppressed by DNA methylation. To directly test whether retroviral repression was due to de novo methylation of the newly integrated retroviruses, we treated the long-term cultures with 5-azadC, a drug that leads to hypomethylation of genomic DNA (16). If DNA methylation was preventing expression of the MSCV LTR, treatment with the drug should activate retroviral expression. Indeed, we found that 5-azadC treatment of ES cells that had lost expression of GFP through long-term passage reactivated the provirus (Fig. 3C). In contrast, *Dnmt1*$^{-/-}$ ES cells infected with the retrovirus did not lose expression of GFP; thus, treatment with 5-azadC did not significantly affect retroviral expression (Fig. 3C). We also analyzed clonal lines containing single proviral integrants in which GFP expression was progressively silenced and found that treatment with 5-azadC resulted in the reactivation of gene expression in all cases (data not shown). This demonstrates that DNA methylation controls long-term but not short-term expression of retroviruses in ES cells.

Expression is maintained after in vitro differentiation. Previously, in vitro differentiation of ES cells had been demonstrated to silence expression of retroviral sequences (12, 20). Thus, we tested whether GFP exprcssion from the MiG retrovirus in ES cells was affected by in vitro differentiation. We cultured MiG-infected wild-type ES cells in the absence of embryonic feeder cells and LIF in suspension to generate embryoid bodies. Disaggregated embryoid bodies were replated either with or without retinoic acid. We found no change in GFP expression in MiG-infected bulk cultures or individual subclones containing one to several integrants upon in vitro differentiation with either method, as shown for one clonal line containing multiple integrants in Fig. 3D. GFP expression was unchanged in all in vitro-differentiated ES cell lines, regardless of whether the subclones contained only a single or multiple integrants. This indicates that the MSCV-based MiG retrovirus is not silenced by in vitro differentiation.

Generation of mice from GFP-expressing MiG-infected ES cells. We next determined whether expression of the MSCV-based MiG vector was affected by in vivo differentiation of the infected ES cells. Cells from the chimeric animals were derived by injection of MiG-infected wild-type ES cells (derived from 129-Sv/Jae mice) into BALB/c blastocysts. MiG-infected *Dnmt1*$^{-/-}$ ES cells cannot be used for injection into blastocysts, because *Dnmt1*$^{-/-}$ ES cells die upon differention and therefore do not contribute significantly to adult mice (22). MiG-infected wild-type ES cells were sorted for GFP expression by flow cytometry prior to injection, and two GFP-expressing clones, R2 and R11, were isolated (Fig. 4B). Southern blot analysis demonstrated that R2 contained two integrants that comigrate on an agarose gel, and R11 contained three proviral integrants (Fig. 4A). High-contribution chimeras (>80% by coat color) were generated from the R2 and R11 ES cells, which transmitted the proviruses to their offspring (data not shown).

To test whether the chimeras expressed the integrated retroviruses, we isolated peripheral blood mononuclear cells

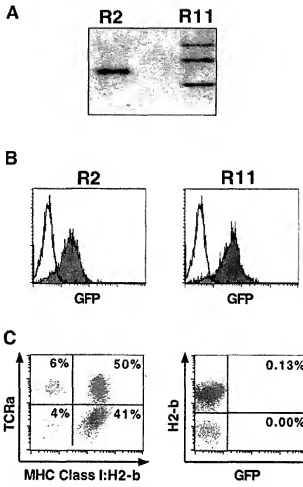


FIG. 4. Retrovirally infected, GFP-expressing ES cells generated nonexpressing mice. (A) Two retrovirally infected clones sorted for GFP expression were analyzed for proviral integrants by Southern blot analysis. R2 contained two integrants, while R11 contained three. Uninfected cells are negative. (B) The clones were passaged after sorting for GFP-expressing cells by flow cytometry and reanalyzed for GFP expression. Both R2 and R11 express GFP (shaded) compared to uninfected controls (unshaded). (C) PBMCs from the R2 chimera (more than 50% contribution by coat color) were analyzed by flow cytometry. ES cell contribution to the chimera was determined by phycoerythrin-H2-b staining and cyc-TCRa staining and demonstrated contribution to the T-cell compartment. The percentage of cells in each quadrant is listed. The cells were also monitored for GFP expression. The percentage of GFP$^+$ cells that are either major histocompatibility complex (MHC) class I H2-b$^+$ (ES cell derived) or H2-b$^-$ (blastocyst derived) is listed in the quadrant.

(PBMCs) from both the R2- and R11-derived chimeras. To distinguish whether the PBMCs were derived from the ES cell donor or the host blastocyst, we stained the cells with antibodies that recognized specific major histocompatibility complex class I haplotypes (Pharmingen). The donor ES cells (129 derived) are H2-b, and the blastocysts (BALB/C derived) are H2-d (Fig. 4C and data not shown). In addition, we stained the PBMCs with a pan-T-cell (TCRa) (Fig. 4C) or pan-B-cell (B220) antibody (data not shown) to determine the ES cell contribution to these lineages. Using this strategy, we found that approximately 90% of the PBMCs from either the R2 or R11 chimera were ES cell derived as measured by H2-b staining (Fig. 4C, data not shown). However, the majority of the cells did not express GFP in either chimera (Fig. 4C and data not shown). On the order of 0.1% of the PBMCs that were ES

Downloaded from mcb.asm.org by on September 14, 2009

cell derived were GFP positive, compared to less than 0.01% that were blastocyst derived (Fig. 4C). Similar results were also obtained with cells from the R11 chimera (data not shown). The results indicate that the MSCV LTR is repressed during in vivo differentiation to lymphocytes. Nevertheless, a small number of cells escaped silencing and expressed GFP. This transcriptional repression of the MiG provirus in the chimeras is in contrast to the GFP expression both in the donor ES and after in vitro differentiation (Fig. 4B, data not shown).

To determine if other somatic cells expressed the retroviral integrants, we analyzed the progeny of the chimeras. We isolated spleen, thymus, kidney, and liver cells from an animal carrying the two proviral integrants present in the R2 chimera and a littermate control containing no retroviral integrants. We analyzed these cells for GFP expression by FACS analysis and found no detectable expression of GFP in the splenocytes, thymocytes, renal cells, or hepatocyes (Fig. 5A and data not shown).

In vitro reactivation of retroviral expression. One possible explanation for transcriptional repression during in vivo differentiation was de novo methylation of the integrated retroviral LTR during embryonic development. To test this hypothesis, we cultured splenocytes from a mouse containing the R2 proviruses and from a littermate control, by treating the cells with either anti-CD3 or anti-CD40 to activate and induce proliferation of the T cells or B cells, respectively (4). We then assayed for GFP expression by flow cytometry and found that proliferation of the splenocytes did not activate expression of the retrovirus (data not shown). Next, we added 5-azadC to the splenocyte cultures to induce demethylation of the retroviruses. Indeed, treatment with 5-azadC activated expression in approximately 2% of the T cells (anti-CD3) (Fig. 5B) and 2% of the B cells (anti-CD40) (data not shown). In addition, when in vivo-differentiated cells, which had been isolated from the kidney of a transgenic mouse and transformed with simian virus 40 large T antigen (15), were treated with 5-azadC, activation of the silent provirus was observed in a similar fraction of the cells (data not shown). The extent of reactivation of expression of the provirus in in vivo-differentiated cells by 5-azadC was lower than in ES cells, where the reactivation of the provirus with 5-azadC was almost complete.

We next determined whether demethylation of the retrovirus in vivo would activate expression of the integrated retroviruses (13). Newborn mice were subcutaneously injected with 5-azadC at postnatal day 5 and subsequently analyzed at postnatal day 14 for GFP expression by flow cytometry. We found that 5-azadC-injected animals but not the uninjected controls had activated GFP expression of the proviruses in the spleen, thymus, and kidney (Fig. 5C and data not shown). When we injected higher concentration of 5-azadC in an effort to further demethylate the newborn mice, all injected animals died. This result demonstrated that repression by DNA methylation is, at least in part, responsible for silencing expression of the retroviral LTR in vivo.

Retroviral expression in HS cells after serial adoptive transfers. Bone marrow contains the HS cells that can stably repopulate the hematopoietic system after transfer to lethally irradiated mice. To determine whether HS cells can be effectively transduced and express the MiG retrovirus, we used infected bone marrow cells to reconstitute lethally irradiated mice (Fig. 6). We found that between 30 and 80% of the splenocytes from these primary recipients expressed the retrovirus, as measured by FACS analysis for GFP expression and shown for one representative experiment (Fig. 6A). The MiG virus was expressed in the B-cell, T-cell, and granulocyte compartments, as measured by a pan-B cell (B220), pan-T-cell (Thy-1), and pan-

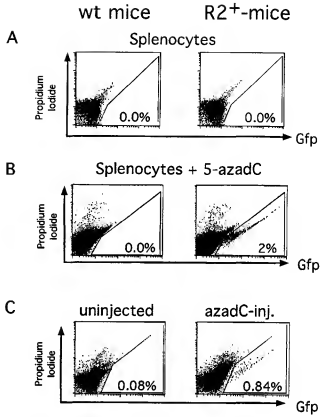


FIG. 5. Silenced retroviruses can be reactivated with 5-azadC. (A) Splenocytes from an $R2^+$ or $R2^-$ littermate do not express GFP by flow cytometry. Cells were stained with propidium iodide to exclude dead cells, and the percentage of GFP^+ cells is indicated. (B) The splenocytes from panel B were induced to proliferate with anti-CD3 and treated with 5-azadC. Cells were stained with propidium iodide to exclude dead cells and analyzed by flow cytometry. The percentage of GFP^+ cells is indicated. (C) Flow cytometric analysis of the splenocytes of littermates that were either uninjected or injected with 5-azadC at passage 5 and analyzed at passage 14 for GFP expression. The percentage of GFP^+ splenocytes is indicated.

granulocyte (Gr-1) marker electronically gated on GFP-positive cells (Fig. 6B and data not shown). Because a large fraction of the splenocytes in the primary recipients are derived from relatively differentiated, lineage-committed progenitors, serial adoptive transfers are required to test for retroviral expression in the true HS cells (17). Therefore, we used bone marrow from these primary recipients to serially reconstitute lethally irradiated mice. This protocol requires substantial expansion from the stem cells and tests for long-term expression of the retrovirus. We observed no change in the percentage GFP-positive HS cells, and the level of GFP expression from the adoptive transfers into multiple recipients was stable over three additional passages (4° recipient). In addition, the infected cells gave rise to both B- and T-cell lineages at the expected ratios (Fig. 6B), demonstrating not only that the MiG retrovirus transduced the long-term repopulating HS cells but also that the MiG-mediated GFP expression was stable during in vivo hematopoietic differentiation. However, our results do not exclude the possibility that in addition to the transcriptionally active proviruses present within these cells, there are also copies of the virus that were transcriptionally silenced.

Downloaded from mcb.asm.org by on September 14, 2009

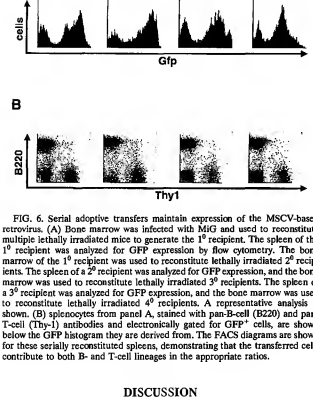


FIG. 6. Serial adoptive transfers maintain expression of the MSCV-based retrovirus. (A) Bone marrow was infected with MiG and used to reconstitute multiple lethally irradiated mice to generate the 1^0 recipient. The spleen of the 1^0 recipient was analyzed for GFP expression by flow cytometry. The bone marrow of the 1^0 recipient was used to reconstitute lethally irradiated 2^0 recipients. The spleen of a 2^0 recipient was analyzed for GFP expression, and the bone marrow was used to reconstitute lethally irradiated 3^0 recipients. The spleen of a 3^0 recipient was analyzed for GFP expression, and the bone marrow was used to reconstitute lethally irradiated 4^0 recipients. A representative analysis is shown. (B) splenocytes from panel A, stained with pan-B-cell (B220) and pan-T-cell (Thy-1) antibodies and electronically gated for GFP^+ cells, are shown below the GFP histogram they are derived from. The FACS diagrams are shown for these serially reconstituted spleens, demonstrating that the transferred cells contribute to both B- and T-cell lineages in the appropriate ratios.

DISCUSSION

We have investigated the role of DNA methylation in retroviral silencing. Retrovirus-based studies of stem cells have been hampered by the lack of expression. We have overcome the transcriptional repression in ES cells by using an MSCV-based vector in combination with a sensitive GFP reporter gene (MiG vector). The analysis of expression of the MiG vector and other Moloney virus-based vectors in *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ ES cells has allowed us to determine whether DNA methylation directly controls retroviral gene expression in these cells. We found that both methylation-dependent and methylation-independent mechanisms exist to control retroviral gene expression.

Historically, retroviral expression of Moloney virus-based vectors in ES cells has been negligible. In contrast, the MSCV LTR not only transduces GFP expression in ES cells but also expresses other exogenous gene products such as the Cre recombinase and the antiapoptotic factor Bcl-2 at detectable level in ES cells. Therefore, the MSCV LTR can be used to express various transgenes in ES cells and their differentiated descendant cells.

It had been proposed that DNA methylation has evolved as a cellular mechanism to silence retroviral elements, preventing the spread of transposable elements through the genome (39). Indeed, de novo methylation of integrated proviral sequences has been observed in wild-type ES cells, which was correlated with the transcriptional silencing of the retrovirus (14). Our findings are the first demonstration that inhibition of the *Dnmt1* methyltransferase gene prevents silencing of the retroviruses in ES cells. This result provides direct evidence that DNA methylation is causally involved in long-term retroviral repression. Consistent with this conclusion is the demonstration that the transcriptionally silenced proviruses present in long-term *Dnmt1*$^{+/+}$ ES cell cultures can be reversed by treatments with 5-azadC.

In contrast, methylation-independent mechanisms determine initial retroviral expression in ES cells. Wild-type or *Dnmt1*$^{-/-}$ ES cells infected with Moloney virus-based vectors were transcriptionally silent, and therefore this silencing was independent of the DNA methylation status of the cells. Moreover, the basal level of expression from the MSCV-based vector was unaffected by the methylation status of the cells. This formally demonstrates that DNA methylation-independent mechanisms control initial retroviral gene expression in ES cells. Because the basal level of expression of the MSCV LTR in ES cells is lower than in differentiated cell types and not affected by the methylation status of the ES cells, *trans*-acting factors must regulate the initial level of expression.

Previous studies found that retroviruses, including the MSCV LTR, are silenced by the in vitro differentiation process (20). In contrast, we found for the first time that expression of this MSCV-based retrovirus in ES cells was maintained after in vitro differentiation with and without retinoic acid. We were also able to show long-term, stable GFP expression from the MiG vector in HS cells and their differentiated derivatives. MiG-mediated GFP expression from HS cells was stable through serial adoptive transfers, and the HS cells gave rise to GFP-expressing B- and T-cell lineages. Therefore, this MSCV-based retroviral transduction system should allow for a molecular analysis of stem cell biology and differentiation programs by forced expression of exogenous gene products.

It has been postulated that methylation-dependent mechanisms repress retroviral gene expression upon in vivo differentiation (13, 20). To test this, we injected GFP-expressing undifferentiated ES cells into recipient blastocysts and generated chimeric mice. Differentiated tissues derived from these in vivo-differentiated ES cells, such as PBMCs, lacked significant GFP expression. Treatment of ES cell-derived differentiated cells with 5-azadC in vitro or in vivo led to partial reactivation of expression of the silenced retroviruses in lymphoid and nonlymphoid tissues. We conclude from these results that the maintenance of retroviral silencing in vivo involves DNA methylation. However, only a small fraction of the 5-azadC-treated cells reactivated GFP expression, unlike the long-term ES cell cultures, in which every cell reactivated GFP expression. This suggests that methylation-independent mechanisms exist to suppress retroviral expression. Alternatively, 5-azadC treatment of differentiated cells, in contrast to ES cells, may not lead to a level of genomic demethylation sufficient for complete retroviral reactivation. The transgenic animals carrying the silenced MiG proviruses will be a valuable indicator for in vivo activation of GFP expression under different conditions.

ACKNOWLEDGMENTS

The first two authors, S. R. Cherry and D. Biniszkiewicz, contributed equally to this work.

We thank members of the Baltimore and Jaenisch labs for advice and discussions. We are grateful to Brian Bates for discussions and comments on the manuscript, George Daley for discussions, Jessie Dausman for blastocyst injections, Ruth Curry for mouse work, and Glen Paradis for FACS sorting.

D. Biniszkiewicz was supported by the Deutsche Akademische Austauschdienst. D. Baltimore and R. Jaenisch (NIH/NCI 5-R35-CA44339) are supported by NIH grants. This work was supported in part by the ERC program of the National Science Foundation under award EEC-9843342.

REFERENCES

1. **Asche, W., G. Colletta, G. Warnecke, P. Nobis, S. Pennie, R. M. King, and W. Ostertag.** 1984. Lack of retrovirus gene expression in somatic cell hybrids of

Downloaded from mcb.asm.org by on September 14, 2009

Friend cells and teratocarcinoma cells with a teratocarcinoma phenotype. Mol. Cell. Biol. 4:923–930.
2. **Cedar, H.** 1988. DNA methylation and gene activity. Cell 53:3–4.
3. **Challita, P. M., and D. B. Kohn.** 1994. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. Proc. Natl. Acad. Sci. USA 91:2567–2571.
4. **Coligan, J. E., et al. (ed.).** Current protocols in immunology. J. Wiley & Sons, New York, N.Y.
5. **Gautsch, J. W.** 1980. Embryonal carcinoma stem cells lack a function required for virus replication. Nature 285:110–112.
6. **Gautsch, J. W., and M. C. Wilson.** 1983. Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature 301:32–37.
7. **Greiser-Wilke, I., W. Ostertag, P. Goldfarb, A. Lang, M. Furusawa, and J. F. Conscience.** 1981. Inducibility of spleen focus-forming virus by BrdUrd is controlled by the differentiated state of the cell. Proc. Natl. Acad. Sci. USA 78:2995–2999.
8. **Grez, M., E. Akgun, F. Hilberg, and W. Ostertag.** 1990. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. Proc. Natl. Acad. Sci. USA 87:9202–9206.
9. **Hoeben, R. C., A. A. Migchielsen, R. C. van der Jagt, H. van Ormondt, and A. J. van der Eb.** 1991. Inactivation of the Moloney murine leukemia virus long terminal repeat in murine fibroblast cell lines is associated with methylation and dependent on its chromosomal position. J. Virol. 65:904–912.
10. **Huang, D. C., S. Cory, and A. Strasser.** 1997. Bcl-2, Bcl-XL and adenovirus protein E1B19kD are functionally equivalent in their ability to inhibit cell death. Oncogene 14:405–414.
11. **Jacob, J., and D. Baltimore.** 1999. Modelling T-cell memory by genetic marking of memory T cells in vivo. Nature 399:593–597.
12. **Jaenisch, R., H. Fan, and B. Croker.** 1975. Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. Proc. Natl. Acad. Sci. USA 72:4008–4012.
13. **Jaenisch, R., A. Schnieke, and K. Harbers.** 1985. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. Proc. Natl. Acad. Sci. USA 82:1451–1455.
14. **Jahner, D., and R. Jaenisch.** 1985. Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. Nature 315:594–597.
15. **Jat, P. S., C. I. Cepko, R. C. Mulligan, and P. A. Sharp.** 1986. Recombinant retroviruses encoding simian virus 40 large T antigen and polyomavirus large and middle T antigens. Mol. Cell. Biol. 6:1204–1217.
16. **Jones, P. A.** 1985. Altering gene expression with 5-azacytidine. Cell 40:485–486.
17. **Jones, R. J., J. E. Wagner, P. Celano, M. S. Zicha, and S. J. Sharkis.** 1990. Separation of pluripotent haematopoietic stem cells from spleen colony-forming cells. Nature 347:188–189.
18. **Kohn, D. B.** 1995. The current status of gene therapy using hematopoietic stem cells. Curr. Opin. Pediatr. 7:56–63.
19. **Krall, W., and D. B. Kohn.** 1996. Expression levels by retroviral vectors based upon the N2 and the MFG backbones. Gene Ther. 3:365.
20. **Laker, C., J. Meyer, A. Schopen, J. Friel, C. Heberlein, W. Ostertag, and C. Stocking.** 1998. Host *cis*-mediated extinction of a retrovirus permissive for expression in embryonal stem cells during differentiation. J. Virol. 72:339–348.
21. **Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li.** 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122:3195–3205.
22. **Li, E., T. H. Bestor, and R. Jaenisch.** 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926.
23. **Loh, T. P., L. L. Sievert, and R. W. Scott.** 1990. Evidence for a stem cell-specific repressor of Moloney murine leukemia virus expression in embryonal carcinoma cells. Mol. Cell Biol. 10:4045–4057.
24. **Naviaux, R. K., E. Costanzi, M. Haas, and I. M. Verma.** 1996. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. J. Virol. 70:5701–5705.
25. **Nichols, J., E. P. Evans, and A. G. Smith.** 1990. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. Development 110:1341–1348.
26. **Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara.** 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell 32:1105–1113.
27. **Onishi, M., S. Kinoshita, Y. Morikawa, A. Shibuya, J. Phillips, L. L. Lanier, D. M. Gorman, G. P. Nolan, A. Miyajima, and T. Kitamura.** 1996. Applications of retrovirus-mediated expression cloning. Exp. Hematol. 24:324–329.
28. **Pear, W. S., G. P. Nolan, M. L. Scott, and D. Baltimore.** 1993. Production of high-titer helper-free retroviruses by transient transfection. Proc. Natl. Acad. Sci. USA 90:8392–8396.
29. **Petersen, R., G. Kempler, and E. Barklis.** 1991. A stem cell-specific silencer in the primer-binding site of a retrovirus. Mol. Cell. Biol. 11:1214–1221.
30. **Prasad, V. E., and P. W. Rigby.** 1991. Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. J. Virol. 65:1803–1811.
31. **Robbins, P. B., D. C. Skelton, X. J. Yu, S. Halene, E. H. Leonard, and D. B. Kohn.** 1998. Consistent, persistent expression from modified retroviral vectors in murine hematopoietic stem cells. Proc. Natl. Acad. Sci. USA 95:10182–10187.
32. **Robertson, E., A. Bradley, M. Kuehn, and M. Evans.** 1986. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323:445–448.
33. **Silver, D. P., E. Spanopoulou, R. C. Mulligan, and D. Baltimore.** 1993. Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. Proc. Natl. Acad. Sci. USA 90:6100–6104.
34. **Soriano, P.** 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. 21:70–71.
35. **Tsukiyama, T., H. Ueda, S. Hirose, and O. Niwa.** 1992. Embryonal long terminal repeat-binding protein is a murine homolog of PTZ-F1, a member of the steroid receptor superfamily. Mol. Cell. Biol. 12:1286–1291.
36. **Van Parijs, L., Y. Refaeli, A. K. Abbas, and D. Baltimore.** 1999. Autoimmunity as a consequence of retrovirus-mediated expression of C-FLIP in lymphocytes. Immunity 11:763–770.
37. **Van Parijs, L., Y. Refaeli, J. D. Lord, B. H. Nelson, A. K. Abbas, and D. Baltimore.** 1999. Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. Immunity 11:281–288.
38. **Walsh, C. P., J. R. Chaillet, and T. H. Bestor.** 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. 20:116–117.
39. **Yoder, J. A., C. P. Walsh, and T. H. Bestor.** 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13:335–340.

Downloaded from mcb.asm.org by on September 14, 2009

TAB 6

This material may be protected by Copyright law (Title 17 U.S. Code)

Clin Genet 2001: 59: 17–24
Printed in Ireland. All rights reserved

Copyright © Munksgaard 2001
CLINICAL GENETICS
ISSN 0009-9163

Developmental Biology: Frontiers for Clinical Genetics

Section Editor: Roderick R McInnes
e-mail: mcinnes@sickkids.on.ca

The beta-globin locus control region versus gene therapy vectors: a struggle for expression

Ellis J, Pannell D. The beta-globin locus control region versus gene therapy vectors: a struggle for expression.
Clin Genet 2001: 59: 17–24. © Munksgaard, 2001

Developmental control of gene expression has a major impact on the design of β-globin retrovirus vectors for hematopoietic stem cell gene therapy of β-thalassemia. It is obvious that the endogenous locus control region (LCR) elements that drive β-globin gene expression in transgenic mice must be included in these vectors. However, the specific elements to use are not clear and require an understanding of LCR action. Moreover, retrovirus vectors contain silencer elements that function in stem cells and are dominant to LCR function. Recent studies on LCRβ-globin transgenes and retrovirus silencing suggest ways to overcome this silencing effect after transfer into stem cells and carefully designed lentivirus vectors have exciting therapeutic benefit in animal models of β-thalassemia. By building on 15 years of development, LCRβ-globin vectors are now being tested in preclinical animal models and may ultimately lead to the long-sought cure for this genetic disease.

J Ellis and D Pannell

Developmental Biology Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON, and Department of Molecular and Medical Genetics, University of Toronto, Canada.

Key words: chromatin – gene silencing – gene therapy – LCR

Corresponding author: J Ellis, Developmental Biology Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON, Canada M5G 1X8. Tel: +416 813 7295; fax: +416 813 8833; e-mail: jellis@sickkids.on.ca

Received 12 October 2000, revised and accepted for publication 16 October 2000

β-thalassemia is caused by genetic defects that reduce β-globin protein levels (1). The resulting imbalance of β-globin to α-globin chains results in a severe anemia that is generally treated by repeated blood transfusions. A consequence of such regular blood transfusion is the increased risk of exposure to virus-infected blood supplies, and eventually leads to iron overload that can cause organ failure (2). Present iron chelation treatments are controversial (3), and the only cure available is bone marrow transplantation from a matched sibling if available. An attractive alternative is to perform gene therapy to deliver a human β-globin gene into hematopoietic stem cells (HSC) from the patient (4, 5).

For β-globin gene therapy to be successful, it is essential that the transferred gene be expressed to the correct level. This goal requires a detailed knowledge of the mechanism and *cis*-acting sequences that control β-globin expression during development. Surprisingly, inclusion of the appropriate regulatory elements may not be sufficient to obtain therapeutic levels during gene therapy as the vectors used to deliver the gene are frequently silenced in transduced stem cells. Hence, a better understanding of the mechanism of vector silencing in stem cells is also required. Here, we review the regulatory elements that control β-globin gene expression during development and their use in β-globin retrovirus vectors, outline the evidence that retrovirus and lentivirus vectors are silenced in stem cells and potential means to overcome this silencing, and we conclude with preclinical animal models to test promising β-globin gene therapy vectors.

β-globin gene expression during development

The β-globin gene is part of a cluster of highly related globin genes located on Chr 11p15 in humans (6). These genes are arranged in the same order as they are expressed during development (Fig. 1). The ε-globin gene is expressed in the blood islands of the yolk sac, the site of hemato-

65/264

poiesis then switches to the fetal liver where the γ-globin genes are expressed, and shortly after birth hematopoiesis switches to the bone marrow where the δ- and β-globin genes are expressed to very low and high levels, respectively. Many mutations that cause β-thalassemia have been described and some have been informative with regard to the sequences and molecular mechanisms that control globin gene switching. For example, it is clear that point mutations in the γ-globin promoters can enhance expression of these genes in adults causing hereditary persistence of fetal hemoglobin (HPFH). Deletion of upstream sequences that include the locus control region (LCR) results in a lack of expression from the still intact globin genes. These data demonstrate that promoter sequences and the LCR are important control elements for globin expression.

The β-globin locus control region

The LCR is composed of at least four DNaseI hypersensitive sites (HS) located upstream of the locus (Fig. 1) (7–9). The presence of HS indicates that *trans*-acting factors are binding to these regions and displacing or destabilizing nucleosomes. Nucleosomes are the basic units of chromatin and condense DNA around an octomer of the histone proteins H2A, H2B, H3 and H4 (10). Expressed genes are located an 'open' chromatin that is more accessible to *trans*-acting factors and in general contain nucleosomes with highly acetylated histones. In contrast, 'closed' chromatin generally has deacetylated histones and is bound by the linker histone H1, is less accessible to DNA binding factors and genes in these regions are not expressed. Chromatin structure is modulated using many chromatin remodelling complexes (11, 12).

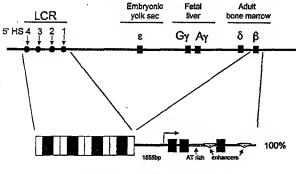


Fig. 1. Gene structure and expression pattern of the human β-globin locus. The locus control region (LCR) is composed of at least 4 DNaseI hypersensitive sites (HS) located upstream of the cluster. *Trans*-acting factors (red circles) bind to the HS. In transgenic mice, LCR activity directs copy number-dependent, position-independent transgene expression. Full levels of β-globin transgene expression are obtained in the presence of all four HS including their core elements (red boxes).

The mechanism by which the LCR controls β-globin gene expression has been extensively studied, primarily using transgenic and knockout mice (6). It is important to realize that these two assays manipulate the genes in quite different ways and the results are not always complementary or in agreement with each other. Transgenic mice contain the human β-globin gene transferred into novel or ectopic integration sites; whereas, knockout mice manipulate the endogenous native locus in the mouse. In transgenic mice, human β-globin transgenes are silent at most integration sites or transcribed to about 1% of the endogenous mouse βmajor level. In contrast, addition of the LCR including all four HS to the β-globin transgene results in expression to about 100% levels at all integration sites, and expression is copy number-dependent (Fig. 1) (7, 13, 14). This copy number-dependent, position-independent transgene expression is unusual and is the defining feature of LCR activity. Further investigation demonstrated that individual HS2, HS3 and HS4 elements and their smaller 'cores' of approximately 200–300 bp, also direct copy number-dependent transgene expression but to lower levels (10–25%) (15–20).

The LCR is often referred to as an enhancer, but does not have classic enhancer activity because it does not function equally well in either orientation (21). Rather, it appears that complete LCR activity requires all four HS (22), and these have some distinct roles. For example, HS3 can activate β-globin transgenes at all single copy integration sites where it establishes open chromatin and remodels chromatin on the promoter to permit expression (13). In contrast, although HS2 has strong enhancer activity in transient transfection studies (23), it is unable to direct expression in single copy transgenic mice (24). These data suggest that at ectopic sites, the HS function together as a unit, making the LCR sufficient to open chromatin and enhance full expression of β-globin transgenes.

Open chromatin is likely to be established by the binding of erythroid *trans*-acting factors that recruit chromatin remodelling complexes (25–31), as has been described for histone acetylation changes on active human β-globin genes (32). This open chromatin may not extend throughout transgenes containing the entire human LCRβ-globin cluster, as different domains that correlate with the presence of low level intergenic transcription have been described during globin switching in mice (33).

Two models of LCR activity

The transgenic mouse data have largely been interpreted as supporting a holocomplex model of LCR

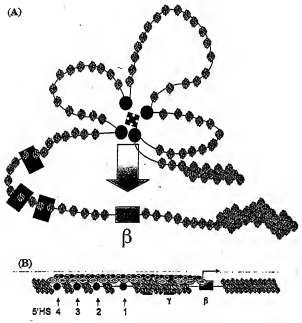


Fig. 2. Two models of LCR activity. A) The holocomplex model suggests that factors (red circles) bound to the HS interact with each other by DNA looping (arrows) to form a single LCR holocomplex that loops to activate expression from the appropriate globin promoter. B) The linking model suggests that the LCR serves to ensure that factors (coloured circles) are bound throughout the locus. Nucleosomes in 'open' chromatin are represented as dispersed green circles, 'closed' chromatin as condensed solenoids of green circles.

action at ectopic sites (34). Once open chromatin has been established, it is proposed that each of the HS then interacts with each other by DNA looping mediated via the bound factors to form an LCR holocomplex (Fig. 2A). The holocomplex would then interact with a single globin gene in the cluster, and switching during development would be accomplished by stage-specific silencer elements associated with the ε- and γ-globin genes. Although there is no direct physical evidence for DNA looping between the LCR and the globin promoters *in vivo*, the holocomplex model is supported by findings that only a single globin gene is transcribed at a time in transgenic mice containing the whole human LCRβ-globin cluster (35), and that the LCR preferentially activates genes closest to it (36). Deletions that remove only a single HS 'core' element drastically reduce transgene expression in comparison to those that delete an entire HS fragment (37, 38), indicating that removal of a core creates a defective holocomplex (39).

Quite different conclusions have been arrived at using knockout technology on the mouse β-globin locus (40). Deletions of individual or all the HS in the endogenous locus do not alter chromatin structure and have relatively minor effects on expression of the globin genes (41, 42). These data suggest that the LCR is not required for chromatin opening at the endogenous mouse β-globin locus, and suggest that more distant elements control chromatin structure (43). A linking model that does not invoke DNA looping has been proposed to explain the knockout results (Fig. 2B) (44). In this model, the function of the LCR is to enhance β-globin expression by ensuring that factors are bound at intervals across the cluster and that the gene is localized to the right nuclear compartment (32). The linking model is not consistent with the ability of the LCR to open chromatin at ectopic transgene sites, but the holocomplex model cannot easily explain the effect of LCR deletions in the mouse β-globin locus. As described below, the two models are not necessarily mutually exclusive and may be strengthened by being merged. For the purpose of gene therapy where globin expression cassettes delivered by viral vectors must express at ectopic sites, it will be important to design the cassettes based on transgene constructs that express to high levels at single copy integration sites.

LCRβ-globin expression cassettes for gene therapy

To express therapeutic levels of β-globin from gene therapy cassettes, full expression levels should be obtained from a single copy integration in order to convert a null thalassemia into an asymptomatic carrier state. Initial β-globin gene therapy cassettes were designed to be as small as possible to facilitate gene transfer and used either cDNA or genomic β-globin fragments controlled by minimal promoters (45–47). Addition of small HS core elements improves expression in tissue culture experiments, but were largely disappointing when transferred into mouse bone marrow cells (48, 49). Recent single copy transgenic mouse experiments now demonstrate that full expression by the LCR requires all four HS and specific elements within the β-globin gene including the −1555 bp promoter and the 3′ enhancer (Fig. 3 top) (50). As this 8.8-kb cassette is too large for conventional retrovirus vectors, smaller constructs that express highly are required for retrovirus delivery. A very promising new 3.9-kb cassette expresses γ-globin mRNA to 70% levels in single copy transgenic mice, using β-globin promoter, intron 2 and 3′ enhancer elements (Fig. 3 centre) (51). This construct, or others shown to function effectively in transgenic mice, may ultimately prove to be best suited for gene therapy and take advantage of the anti-sickling properties of γ-globin. In addition, these data demonstrate that the LCR must functionally interact with more than just the promoter. The simplest

interpretation merges the holocomplex and linking models by suggesting that the LCR loops to interact with the promoter, but factors must also be bound throughout the gene.

Retrovirus and lentivirus vectors

Retrovirus vectors have been the method of choice for delivering β-globin genes into hematopoietic stem cells because they stably integrate at single copy into the genome (52). In practice, it has been very difficult to obtain high titer β-globin retrovirus vectors due to instability of the LCR elements and globin intron 2 sequences (48, 49). To stabilize transmission of intact β-globin genes it is necessary to use only certain combinations of HS sites and to delete an AT-rich region in intron 2. Although these modifications permit generation of high titer retrovirus, it is now apparent that the deleted AT-rich sequence is required for high level expression (51).

A limitation of retrovirus vectors is that they integrate only into cycling cells, and the target HSC for β-globin gene therapy tend to be non-cycling. In this regard, lentivirus vectors based on HIV-1 (53) are far superior as they integrate into non-cycling cells and contain the RRE element that is bound by the Rev protein to stabilize the virus genomic RNA (52). One exciting report of a β-globin lentivirus vector demonstrates that it is possible to transmit large LCR fragments coupled to a β-globin cassette with a small promoter, AT-rich deleted intron 2 and no 3′ enhancer (Fig. 3 bottom). Therapeutic levels of β-globin mRNA and protein were shown in transduced bone marrow in a mouse model of β-thalassemia (54).

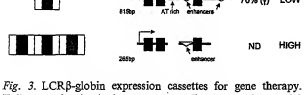


Fig. 3. LCRβ-globin expression cassettes for gene therapy. Full expression in single copy transgenic mice requires an 8.8 kb construct (top) including all four HS, the large β-globin promoter, β-globin intron 2 AT-rich region, and the intron 2 and 3′ enhancers (yellow triangles). High level expression of γ-globin exons (Blue boxes) is obtained from a smaller 3.9 kb β/γ-globin hybrid cassette (centre). As both of these cassettes are not stable in retrovirus vectors, cassettes with improved stability (bottom) have deletions of deleterious sequences.

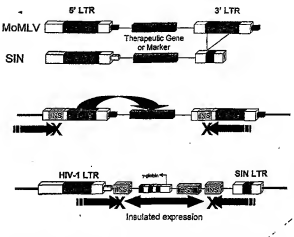


Fig. 4. Retrovirus silencing in stem cells. Silencer elements in MoMLV vectors are primarily located in the LTRs (top), and silencing can be ameliorated but not eliminated by SIN mutations in the 3′LTR. Insulator elements (INS) in the LTR (centre) block silencing by position effects (red arrows with X) in MEL cells but not retrovirus silencing in stem cells (large red arrow). The ideal vector (bottom) uses a SIN lentivirus vector, insulator elements flanking a small HS3β/γ-globin expression cassette to prevent silencing (green arrow), and no marker gene.

Silencing of retrovirus vectors in stem cells

Once the β-globin expression cassette has been transduced into an HSC, it must express to appropriate levels in erythroid cells produced from the HSC. One concern that has severely limited clinical gene therapy is that retrovirus-transduced genes are often silenced in HSC (52, 55–57). It has long been known that retroviruses are silenced in embryonic stem cells and in transgenic mice (58, 59). Moreover, it has been shown that retrovirus and lentivirus sequences silence LCRβ-globin transgenes in mice (60, 61), and that the transgenes fail to open chromatin (60). These data demonstrate that the vectors have silencer elements, and suggest that they are recognized by specific factors that are restricted to stem cells in mammals.

There are several possible means to overcome vector silencing in stem cells. First, the silencer elements can be defined in the retrovirus or lentivirus vectors and, if not required for virus replication, can be removed. As the silencers tend to be located in the viral long terminal repeats (LTRs) that control virus transcription (Fig. 4 top), mutations produce self-inactivating (SIN) virus that is inherently safer. Several SIN retrovirus vectors have been developed from Moloney murine leukemia virus that express to higher levels in ES cells and transgenic mice (62–65). Nevertheless, none of the SIN vectors express at all integration sites in stem cells. To overcome this residual silencing in the SIN vectors, it may be possible to

incorporate insulator elements to block the spread of silencing from surrounding retrovirus sequences (66). The chicken β-globin cHS4 element is an insulator (66, 67) and has been placed in the LTR of retrovirus vectors to block position effects in murine erythroleukemia (MEL) cells (Fig. 4 centre) (68, 69). However, an insulator in the LTR does not shield internal genes from silencing established on the retrovirus sequences in ES cells (68). A better construct design would position insulators on both sides of the internal gene rather than in the LTR (Fig. 4 centre).

Mechanism of retrovirus vector silencing

An understanding of the mechanism of retrovirus silencing may permit interventions that prevent its establishment or maintenance (70, 71). Many groups have correlated *de novo* cytosine methylation (72) of CpG dinucleotides in retroviruses with the silenced state (58, 59, 68, 73, 74), and expression can be reactivated to a low level using the methylation inhibitor 5 AzaCytidine (5AzaC) (75). However, this indirect evidence does not address whether methylation is a cause of silencing or a consequence. Recent direct evidence that retrovirus silencing is independent of *de novo* methylase function has been obtained using dnmt3 knockout ES cells and transgenic *Drosophila* that have no methylase activity (60). As chromatin of retrovirus-silenced LCRβ-globin transgenes in mice is inaccessible to DNaseI and marked by deacetylated histone H3 and bound linker H1 (60), it appears that chromatin modifications play a role in retrovirus silencing. Attempts to relieve silencing may, therefore, require inhibitors of histone deacetylases (HDAC) or of H1 binding.

Silencing is often established by one pathway and maintained by another. Time course experiments in ES cells demonstrate that most retroviruses are silenced within 2 days but some integration sites escape complete silencing and express to low levels (60, 76, 77). As methylation is not detectable by 2 days post-infection, methylation is likely to be a consequence of, or secondary step in, retrovirus silencing. Most experiments have focussed on the subset of infected cells that initially express. Over time, these are gradually silenced in a process known as extinction. In infected MEL cells, extinction can be reversed early in the process using the HDAC inhibitor Trichostatin A (TSA) (78). However, the methylation inhibitor 5 AzaC is required in addition to TSA to overcome extinction at later time points. These data suggest that methylation is an important secondary or associated step in extinction of virus expression in mature cell types. The ability of TSA or 5 AzaC to activate expression in the majority of transduced cells that are completely silenced from the outset has not been rigourously tested to date. In summary, efforts to prevent retrovirus silencing using methylation and HDAC inhibitors hold promise, but require more knowledge of the mechanism and demonstration of their utility in silenced stem cell populations prior to extinction.

An ideal LCRβ-globin lentivirus vector

The ideal β-globin gene therapy vector should stably integrate into an HSC at high efficiency and be expressed to near endogenous levels at single copy. To accomplish this goal, lentivirus vectors have a clear advantage in their ability to infect non-cycling stem cells and stably transmit large LCRβ-globin expression cassettes. The best existing LCRβ-globin lentivirus vector has these features (54), with the additional advantage of omitting a selectable marker gene. Although marker genes are convenient for determining transduction frequencies, most are derived from non-mammalian sources and may themselves be subject to gene silencing effects. However, the vector can be optimized further to direct expression at all integration sites and for vector safety (Fig. 4 bottom). First, a SIN version of the lentiviral vectors must now be used with third generation packaging systems designed to prevent recombination events that generate replication-competent HIV-1 virus (79). The SIN lentivirus vector will not only prevent rescue and spread of the vector by any helper virus, but may also improve expression of the LCRβ-globin cassette (60). Second, the LCRβ-globin cassette should be flanked by insulator elements that are known to block silencing. In practise, this may require a different insulator on one side than the other to avoid recombination events that delete the LCRβ-globin cassette. Finally, the LCRβ-globin cassette should express highly at single copy in transgenic mice. Such constructs use HS3 coupled to a large β-globin promoter, the β-globin intron 2 including both the AT-rich region and enhancer and the 3′ enhancer (51). Use of hybrid genes permits expression of anti-sickling γ- or δ-globin coding sequences instead of β-globin exons (51, 80). A combination of these components should create a lentivirus vector that is safe and expresses therapeutic levels of globin.

Preclinical models of β-thalassemia

Promising β-globin gene therapy vectors have been tested primarily in MEL cells, or in infected mouse

bone marrow. Expression in erythroid cells derived from primitive progenitor cells can be assayed using CFU-S assays in which spleen colonies are formed after 12 days *in vivo*. Expression in cells derived from an infected HSC must be assayed after long-term repopulation assays, followed by secondary transplantation. To show therapeutic efficacy, these long-term studies should be performed first on mouse models of β-thalassemia and sickle cell anemia. A variety of these models have been created by gene targeting and transgenic technology (81–83). These mouse models can be corrected by expression of γ-globin transgenes, and it has been shown that an LCRβ-globin lentivirus vector can express β-globin mRNA and protein in these models with therapeutic benefit (54).

Despite success with the mouse models, it has proved much more difficult to transduce human HSC than murine HSC (84). Fortunately, another preclinical model of β-globin gene therapy into human stem cells is available (Fig. 5). The NOD-Scid mouse is severely immunocompromized and fails to reject human bone marrow transplants (85). Transplanted human HSC home to the mouse bone marrow where they are supported by the hematopoietic microenvironment. Human bone marrow cells from β-thalassemia and sickle cell patients have been shown to repopulate the bone marrow of these mice and generate human red cells that mimic the disease. This system is very well suited to test expression from LCRβ-globin lentivirus vectors in human stem cells. To this end, patient bone marrow or sorted HSC would be infected with the lentivector *in vitro* prior to *in vivo* assay in the NOD/Scid mice. Hence, long-term expression in human stem cells can be assayed *in vivo* without exposing patients to experimental lentivirus vectors.

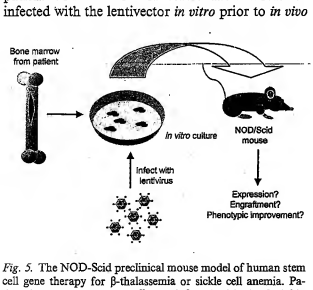


Fig. 5. The NOD-Scid preclinical mouse model of human stem cell gene therapy for β-thalassemia or sickle cell anemia. Patient bone marrow stem cells are infected with the β-globin lentivirus vector *in vitro*, prior to *in vivo* engraftment into immunocompromised NOD-Scid mice. Gene transfer into engrafted human stem cells, β-globin transgene expression, and phenotypic improvement can then be monitored without exposing patients to experimental lentivirus vectors.

Future prospects

The first retrovirus vectors for β-globin gene therapy were designed over 15 years ago, and through slow careful research many obstacles were discovered and gradually surmounted. The success of a well-designed LCRβ-globin lentivirus vector in correcting a mouse β-thalassemia model is a milestone in this process that can now be completed through incremental improvements to vector expression and safety. Ultimately, validation of β-globin gene therapy in human stem cells using the NOD-Scid preclinical model will justify clinical trials of this exciting potential cure for hemoglobinopathies.

Acknowledgements

We acknowledge the Medical Research Council (MRC) of Canada grant support (to JE) for our β-globin expression and retrovirus silencing research and Hospital for Sick Children Foundation, OGS and MRC Doctoral Research Awards to DP. We regret that many fine studies relevant to this review have been omitted due to space constraints.

References

1. Weatherall DJ. The Thalassemias. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001: 183–226.
2. Olivieri NF. The beta-thalassemias. N Engl J Med 1999: 341 (2): 99–109.
3. Olivieri NF et al. Long-term safety and effectiveness of iron-chelation therapy with deferiprone for thalassemia major. N Engl J Med 1998: 339 (7): 417–423.
4. Persons DA, Nienhuis AW. Gene therapy for the hemoglobin disorders: past, present, and future. Proc Natl Acad Sci USA 2000: 97 (10): 5022–5024.
5. Sorrentino BP, Nienhuis AW. Gene therapy for hematopoietic diseases. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. W.B. Saunders: Philadelphia, 2001: 969–1003.
6. Stamatoyannopoulos G, Grosveld F. Hemoglobin switching. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001: 135–182.
7. Grosveld F et al. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 1987: 51 (6): 975–985.
8. Forrester WC et al. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucl Acid Res 1987: 15 (24): 10159–10177.
9. Tuan D, London IM. Mapping of DNase I-hypersensitive sites in the upstream DNA of human embryonic epsilon-globin gene in K562 leukemia cells. Proc Nalt Acad Sci USA 1984: 81 (9): 2718–2722.
10. Felsenfeld G. Chromatin unfolds. Cell 1996: 86 (1): 13–19.

70/264

11. Vignali M et al. ATP-dependent chromatin-remodeling complexes. Mol Cell Biol 2000: 20 (6): 1899–1910.
12. Kingston RE, Narlikar GJ. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev 1999: 13 (18): 2339–2352.
13. Ellis J et al. A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. EMBO J 1996: 15 (3): 562–568.
14. Talbot D et al. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. Nature 1989: 338 (6213): 352–355.
15. Philipsen S et al. The beta-globin dominant control region: hypersensitive site 2. EMBO J 1990: 9 (7): 2159–2167.
16. Philipsen S, Pruzina S, Grosveld F. The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. EMBO J 1993: 12 (3): 1077–1085.
17. Pruzina S et al. Hypersensitive site 4 of the human beta globin locus control region. Nucl Acid Res 1991: 19 (7): 1413–1419.
18. Talbot D et al. Detailed analysis of the site 3 region of the human beta-globin dominant control region. EMBO J 1990: 9 (7): 2169–2177.
19. Talbot D, Grosveld F. The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J 1991: 10 (6): 1391–1398.
20. Caterina JJ et al. Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. Proc Natl Acad Sci USA 1991: 88 (5): 1626–1630.
21. Tanimoto K et al. Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. Nature 1999: 398 (6725): 344–348.
22. Fraser P et al. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev 1993: 7 (1): 106–113.
23. Tuan DY et al. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human 'beta-like globin' genes. Proc Natl Acad Sci USA 1989: 86 (8): 2554–2558.
24. Ellis J et al. Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. EMBO J 1993: 12 (1): 127–134.
25. Blobel GA et al. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. Proc Natl Acad Sci USA 1998: 95 (5): 2061–2066.
26. Boyes J et al. Perturbation of nucleosome structure by the erythroid transcription factor GATA-1. J Mol Biol 1998: 279 (3): 529–544.
27. Boyes J et al. Regulation of activity of the transcription factor GATA-1 by acetylation. Nature 1998: 396 (6711): 594–598.
28. Armstrong JA, Emerson BM. NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 *in vitro*. Mol Cell Biol 1996: 16 (10): 5634–5644.
29. Armstrong JA, Bieker JJ, Emerson BM. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF *in vitro*. Cell 1998: 95 (1): 93–104.
30. Gong QH, McDowell JC, Dean A. Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene *in vivo* by 5' hypersensitive site 2 of the beta-globin locus control region. Mol Cell Biol 1996: 16 (11): 6055–6064.
31. O'Neill D et al. Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. Proc Natl Acad Sci USA 1999: 96 (2): 349–354.
32. Schubeler D et al. Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev 2000: 14 (8): 940–950.
33. Gribnau J et al. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol Cell 2000: 5 (2): 377–386.
34. Grosveld F. Activation by locus control regions? Curr Opin Genet Dev 1999: 9 (2): 152–157.
35. Wijgerde M, Grosveld F, Fraser P. Transcription complex stability and chromatin dynamics *in vivo*. Nature 1995: 377 (6546): 209–213.
36. Dillon N et al. The effect of distance on long-range chromatin interactions. Mol Cell 1997: 1 (1): 131–139.
37. Bungert J et al. Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. Genes Dev 1995: 9 (24): 3083–3096.
38. Peterson KR et al. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. Proc Natl Acad Sci USA 1996: 93 (13): 6605–6609.
39. Wood WG. The complexities of β globin gene regulation. Trends Genet 1996: 12 (6): 204–206.
40. Higgs DR. Do LCRs open chromatin domains? Cell 1998: 95 (3): 299–302.
41. Reik A et al. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. Mol Cell Biol 1998: 18 (10): 5992–6000.
42. Epner E et al. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. Mol Cell 1998: 2 (4): 447–455.
43. Bulger M et al. Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci USA 1999: 96 (9): 5129–5134.
44. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes Dev 1999: 13 (19): 2465–2477.
45. Cone RD et al. Regulated expression of a complete human beta-globin gene encoded by a transmissible retrovirus vector. Mol Cell Biol 1987: 7 (2): 887–897.
46. Dzierzak EA, Papayannopoulou T, Mulligan RC. Lineage-specific expression of a human beta-globin gene in murine bone marrow transplant recipients reconstituted with retrovirus-transduced stem cells. Nature 1988: 331 (6151): 35–41.
47. Karlsson S et al. Expression of the human beta-globin gene following retroviral-mediated transfer into multipotential hematopoietic progenitors of mice. Proc Natl Acad Sci USA 1988: 85 (16): 6062–6066.
48. Sadelain M et al. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. Proc Natl Acad Sci USA 1995: 92 (15): 6728–6732.
49. Leboulch P et al. Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an

TAB 7

This material may be protected by Copyright law (Title 17 U.S. Code)

Repeat-induced gene silencing in mammals

David Garrick[1], Steven Fiering[2,3], David I.K. Martin[2] & Emma Whitelaw[1]

In both plants[1–3] and *Drosophila melanogaster*[4,5], expression from a transgenic locus may be silenced when repeated transgene copies are arranged as a concatameric array. This repeat-induced gene silencing is frequently manifested as a decrease in the proportion of cells that express the transgene, resulting in a variegated pattern of expression. There is also some indication that, in transgenic mammals, the number of transgene copies within an array can exert a repressive influence on expression, with several mouse studies reporting a decrease in the level of expression per copy as copy number increases[6–8]. However, because these studies compare different sites of transgene integration as well as arrays with different numbers of copies, the expression levels observed may be subject to varying position effects as well as the influence of the multicopy array. Here we describe use of the lox/Cre system of site-specific recombination to generate transgenic mouse lines in which different numbers of a transgene are present at the same chromosomal location, thereby eliminating the contribution of position effects and allowing analysis of the effect of copy number alone on transgene silencing. Reduction in copy number results in a marked increase in expression of the transgene and is accompanied by decreased chromatin compaction and decreased methylation at the transgene locus. These findings establish that the presence of multiple homologous copies of a transgene within a concatameric array can have a repressive effect upon gene expression in mammalian systems.

The αPElox construct used to generate transgenic mouse lines is shown in Fig. 1*a*. In this transgene, expression of the *lacZ* reporter gene is driven by the human α-globin promoter and the αHS-40 enhancer-like element of the α-like globin locus[9]. As expression of the *lacZ* reporter can be analysed in single erythroid cells, use of this transgene allows the detection of variegated patterns of expression[10]. The transgene also contains a single copy of the 34-bp recognition site (*loxP*) of the Cre recombinase of bacteriophage P1 (ref. 11). Cre-mediated recombination between *loxP* sites in individual αPElox transgenes that form part of a concatameric array will reduce the transgene copy number without altering the site of integration[12] (Fig. 1*b*). Although transgenes within a multicopy array in mice are usually present in a tandem (head–tail) orientation, inverted repeats (head–head or tail–tail) do occur[13]. Because Cre recombination between inverted *loxP* sites causes the inversion rather than excision of the intervening DNA, only recombination between sites in like-oriented transgene monomers will reduce the copy number of the array.

We generated founder transgenic mice bearing the αPElox construct, and from them we established hemizygous transgenic lines. Southern blotting of tail DNA revealed that two of these transgenic lines (αPElox1 and αPElox7) contain more than 100 copies of the transgene (Fig. 2), and these lines were selected for Cre-mediated reduction in transgene copy number. For each of these parent lines, we collected fertilized oocytes from wild-type female mice mated with G_1 hemizygous transgenic males. Oocytes were micro-injected with a circular Cre expression vector (pCAGGS-Cre) and then transferred into foster mothers. Transient expression of the recombinase from the unincorporated pCAGGS-Cre plasmid has been shown to catalyze efficient site-specific recombination at *loxP* sites within the mouse genome before the morular stage of development[14]. Live-born progeny that showed a reduced transgene copy number as determined by initial Southern blotting were mated with wild-type animals to establish transgenic lines hemizygous for the modified locus. Cre recombination within the αPElox1 parent line (more than 100 copies) gave rise to two distinct reduced-copy progeny lines: 1.cre/a, which contains five copies of the transgene, and 1.cre/b, in which the array has been reduced to a single copy (Fig. 2). Southern-blot analysis indicated that the single copy remaining in the 1.cre/b line contains a rearrangement/deletion (data not shown). The αPElox7 parent line (more than 100 copies) gave rise to a single derivative line (7.cre/a) bearing one copy of the transgene.

For parental and reduced-copy progeny lines, transgene expression was analysed in 12.5-dpc embryos by staining of whole primitive erythrocytes with X-gal. We previously showed that all cells containing β-galactosidase activity can be detected by light microscopy after staining under these conditions[10,15]. For both of the high-copy parent lines, a heavily variegated pattern of transgene expression was observed, with less than 1% of

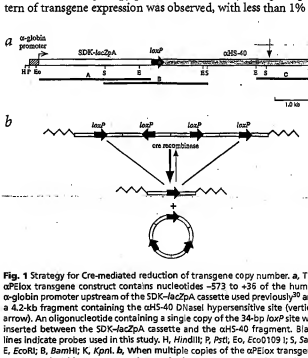


Fig. 1 Strategy for Cre-mediated reduction of transgene copy number. *a*, The αPElox transgene construct contains nucleotides −573 to +36 of the human α-globin promoter upstream of the SDK–*lacZ*pA cassette used previously[30] and a 4.2-kb fragment containing the αHS-40 DNaseI hypersensitive site (vertical arrow). An oligonucleotide containing a single copy of the 34-bp *loxP* site was inserted between the SDK–*lacZ*pA cassette and the αHS-40 fragment. Black lines indicate probes used in this study. H, *Hin*dIII; P, *Pst*I; Eo, *Eco*0109 I; S, *Sac*I; E, *Eco*RI; B, *Bam*HI; K, *Kpn*I. *b*, When multiple copies of the αPElox transgene are situated within a concatameric array in the mouse genome, Cre-mediated recombination between like-oriented *loxP* sites within the array removes the intervening DNA as a circular episome, leaving a reduced-copy array at the same chromosomal location. The forward (excision) reaction is heavily favoured over the reverse (integration) reaction, which requires recombination between *loxP* sites on separate molecules.

[1]*Department of Biochemistry, University of Sydney, New South Wales 2006, Australia.* [2]*Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.* [3]*Current address: Department of Microbiology, Dartmouth Medical Center, Lebanon, New Hampshire, USA. Correspondence should be addressed to E.W. e-mail: e.whitelaw@biochem.usyd.edu.au*

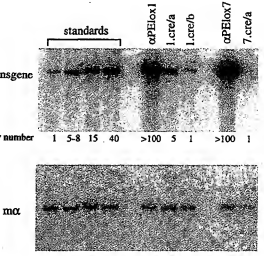


Fig. 2 Copy-number determination for αPElox parental and Cre-modified progeny transgenic lines. DNA was obtained from tail biopsies of three-week-old G_1 hemizygous mice for each of the lines shown, as well as from lines bearing known copy numbers of a transgene that contains the same *lacZ*–αH5-40 cassette. After digestion with *Sac*I, DNA was resolved on a 0.8% agarose gel and transferred to a nylon membrane. The membrane was hybridized with fragment B (Fig. 1a), and the copy number was determined by quantifying the intensity of the resulting signal and comparing it with those of the standards. To ensure equivalent DNA loadings in each lane, the membrane was stripped and re-hybridized with a probe for the murine α-globin gene (mα). In a separate Southern blot performed on tail DNA from each of these lines, hybridization with a probe specific for the 3′ end of the transgene (fragment C, Fig. 1a) revealed that the junction fragments between insert and mouse DNA generated with the restriction enzymes *Eco*RI or *Eco*RV for each of the reduced-copy progeny lines were the same as those observed in the respective parent arrays, indicating that the genomic location of the array had not been altered during Cre modification (data not shown).

erythroid cells containing an active transgene locus (Fig. 3, Table 1). Similar findings of variegated expression patterns in high-copy transgenic lines bearing a construct (αPE) that differs from αPElox only in the absence of the single *loxP* element[16] suggest that the presence of the 34-bp *loxP* oligonucleotide has no effect on transgene expression. When the number of transgene copies within the αPElox1 array was reduced to five by Cre-mediated recombination (1.cre/a), there was a large increase (more than 1,000-fold) in the percentage of primitive erythroid cells expressing the transgene. Similarly, Cre-mediated reduction of the αPElox7 parent array to one copy (7.cre/a) also suppressed the variegated expression that was observed in the parent line, with a 180-fold increase in the size of the expressing population (Fig. 3, Table 1). In these two independent transgenic lines, the presence of high-copy multimeric arrays is therefore associated with a silencing of transgene expression, which is observed as a decrease in the percentage of cells containing an active transgene locus. A decrease in the number of copies within each array correlates with a suppression of variegation—that is, an increase in the size of the expressing population. No expression of the transgene was observed in the line 1.cre/b, in which the single transgene copy was rearranged (data not shown).

In plants, repeat-induced gene silencing has been observed at both the transcriptional[17–19] and post-transcriptional[20,21] stages of gene expression. To determine whether the copy-number–dependent silencing of transgene expression in mouse erythroid cells occurs at transcription or involves a post-transcriptional modification, we performed run-on analysis in nuclei of 12.5-dpc erythrocytes from αPElox1 and its reduced-copy derivative, 1.cre/a (Fig. 4a). Run-on transcripts from the *lacZ* reporter gene were present in 1.cre/a primitive erythrocytes but were not detectable in erythroid cells of the high-copy parent line, indicating that the silencing of expression from multi-copy transgene arrays occurs at the level of transcription. The methylation status and local chromatin structure of the transgene locus in these two lines were also compared. Although transgenes present in 12.5-dpc erythrocytes of the reduced-copy line 1.cre/a appear unmethylated, the high-copy αPElox1 transgene locus at the same genomic location was found to be heavily methylated (Fig. 4b). To analyse chromatin structure, we performed endonuclease protec-

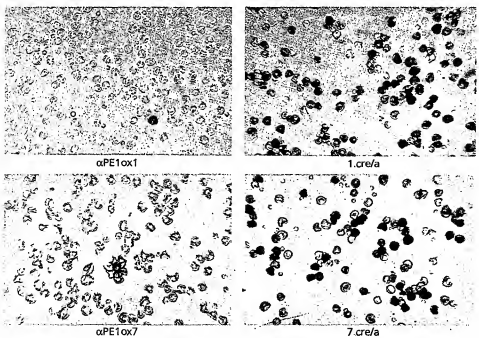


Fig. 3 X-gal staining of primitive erythrocytes from the αPElox transgenic lines and their reduced-copy progeny. Individual 12.5-dpc embryos from the transgenic lines shown were bled into PBS and whole blood cells were stained with X-gal before visualization under light microscopy. For αPElox1, many fields of view had to be scanned to detect a single blue cell.

74/264

Table 1 • Copy number and erythroid expression

Transgenic line	Copy number	Percentage of 12.5-dpc erythroid cells expressing *lacZ*
αPElox1	>100	0–0.01 (n=11)
1.cre/a	5	65±9 (n=11)
αPElox7	>100	0.3±0.02 (n=4)
7.cre/a	1	54±9 (n=3)

Transgene copy numbers were determined by Southern analysis (Fig. 2). The percentage of 12.5-dpc erythroid cells expressing *lacZ* was determined by staining cells as described (Fig.3) and scoring a minimum of 200 cells for a detectable blue colour. Expression data are presented as the mean ± one standard deviation—except for αPElox1, where a range is given. n, number of individual transgenic embryos assayed for each line.

tion assays[22] on nuclei isolated from 12.5-dpc primitive erythroid cells from these two lines. In their native chromatin configuration, transgenes in the high-copy array (αPElox1) were more resistant to endonuclease digestion at a site within the transgene promoter than transgenes that were at the same genomic location but within a lower-copy array (1.cre/a; Fig. 4*c*). The transcriptional silencing that occurs at high-copy arrays is therefore associated both with hypermethylation of transgenes in the locus and with the adoption of a repressive local chromatin configuration.

These results establish that the presence of multiple repeats within a high-copy array can directly repress transgene expression in a mammalian system. Because transgenic animals frequently contain high-copy concatameric arrays of a monomeric unit, this observation has important implications for the meaningful interpretation of transgenic experiments in mammals. The influence of copy number on silencing is likely to be most prominent in lines such as those studied here, in which the array is very large, but repeat-induced silencing may often be responsible for poor transgene expression. In addition to copy number, as shown here, the factors influencing transgene silencing in mice include integration site, the lineage in which expression occurs and the *cis*-acting control elements within the transgene[23]. Transgene constructs for which copy-number–dependent expression has been reported[24,25] may contain genetic elements that function to insulate individual monomers and prevent silencing. Without such elements, the lox/Cre system of site-specific recombination could be used to activate expression in high-copy lines.

The hypermethylation of transgenes within the inactive, high-copy array noted here has also been reported in other cases of repeat-induced silencing[1,3,26,27], although it is unclear in our system whether methylation precedes or is a consequence of the observed chromatin restructuring. The occurrence of copy-number–dependent transgene inactivation in *Drosophila*, in which DNA methylation has not been detected[28], suggests that methylation is not necessary for the chromatin remodelling associated with repeat-induced silencing and that hypermethylation of the αPElox transgene in mice may be a secondary modification that occurs at already inactive, high-copy arrays. This hypothesis is consistent with our previous finding that the variegated silencing of globin/*lacZ* transgenes in mice correlates with an inactive chromatin structure but not hypermethylation at the transgene locus[29].

Although a molecular mechanism for repeat-induced silencing of multimeric arrays is yet to be fully elucidated, the correlation between the silenced state and the adoption of a less accessible

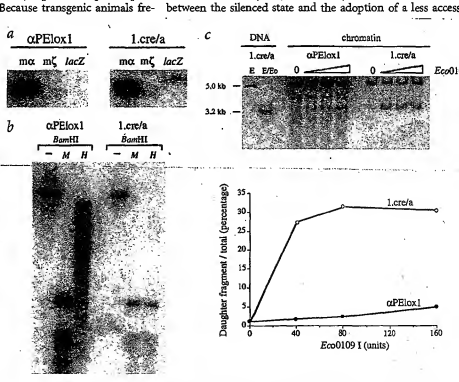


Fig. 4 Transcription, methylation and chromatin accessibility of the transgene. **a**, Nuclear run-on analysis of *lacZ* transcription in αPElox1 and 1.cre/a. Nuclei were prepared from primitive erythroid cells collected from 12.5-dpc transgenic embryos of the lines αPElox1 and 1.cre/a. Run-on transcripts were synthesized and hybridized to membranes containing the following DNA fragments: the mouse α-globin gene from +17 to +956 (mα), the mouse ζ-globin gene from +150 to +969 (mζ) and a fragment of the *lacZ* gene, which includes nucleotides from +39 to +1164 relative to the start site of the αPElox transcript (*lacZ*). **b**, Analysis of methylation status of the transgene locus in αPElox1 and 1.cre/a. DNA from 12.5-dpc erythrocytes of the lines αPElox1 and 1.cre/a was digested with either *Bam*HI alone or *Bam*HI together with either *Msp*I (M) or *Hpa*II (H), which are insensitive and sensitive, respectively, to methylation at the central CpG dinucleotide of their common recognition sequence, 5′–CCGG. The resulting fragments were analysed by Southern blotting and hybridization with fragment A (Fig. 1a). **c**, Endonuclease protection assay of the transgene locus in αPElox1 and 1.cre/a. Nuclei prepared from primitive erythroid cells of the lines αPElox1 or 1.cre/a were digested with increasing concentrations of *Eco*0109 I. The DNA was then purified and digested with *Eco*RI to release a 5.0-kb parent fragment (composed of the tail–head junction between adjacent transgenes) before Southern analysis (top panel). When the membrane was hybridized with fragment A (Fig. 1a), cleavage at the *Eco*0109 I site within the transgene promoter reduced this parent band to a 3.2-kb daughter fragment. The sizes of the expected parent and daughter fragments were determined by digestion of purified tail DNA from the line 1.cre/a with either *Eco*RI alone (E) or both *Eco*RI and *Eco*0109 I (E/Eo). The membrane was quantified with a PhosphorImager and ImageQuant version 4.2a software (Molecular Dynamics; bottom panel). The vertical axis shows the percentage of total signal contained within the daughter fragment at each enzyme concentration for the lines αPElox1 (closed circles) and 1.cre/a (open circles).

chromatin configuration observed both here and in *Arabidopsis*[19] is consistent with a model in which homologous pairing between monomers within the array induces heterochromatinization at the transgene locus. Support for a model of heterochromatin formation is derived from the suppression of copy-number–dependent silencing of pigment genes in the *Drosophila* eye by mutations within genes encoding known structural components of heterochromatin[4]. It remains to be determined whether the repeat-induced modification of chromatin structure in mice is dependent on the proximity of the locus to nearby blocks of constitutive heterochromatin, as was the case for a *brown* eye pigment transgene in *Drosophila*[5], or whether high-copy arrays autonomously form inactive chromatin structures irrespective of their position. The arrangement of endogenous loci such as the rRNA, tRNA and histone genes as high-copy concatamers of a repeated unit suggests that multi-copy arrays need not always be subject to repeat-induced silencing. An investigation of how silencing is prevented at these endogenous loci may prove useful in maintaining activity at multi-copy arrays of foreign genetic elements.

Methods

Transgenic mice. The αPElox transgene was constructed by insertion of an oligonucleotide containing a single copy of the 34 bp *loxP* site (5′–ATAACTTCGTATAATGTATGCTA TACGAAGTTAT–3′) between the SDK–*lacZ*pA cassette and the 4.2-kb αHS-40 fragment of the previously described construct αPE[16]. The integrity of the *loxP* site was confirmed by dideoxy sequencing. The αPElox fragment for micro-injection was excised from plasmid vector sequences by digestion with *HindIII* and *KpnI* and purified by agarose-gel electrophoresis. Transgenic mouse lines bearing the αPElox construct were generated by standard micro-injection techniques in the outbred Pathology Oxford (P.O.) mouse strain. For Cre recombination, fertilized oocytes were collected from wild-type P.O. females mated with hemizygous G_1 transgenic male mice and micro-injected into either pronucleus with the circular Cre-expression vector pCAGGS-Cre[16] at 5 ng/μl before transfer into pseudo-pregnant foster mothers. Live-born progeny that showed a reduced transgene copy number as determined by initial Southern blotting were mated with wild-type animals to establish transgenic lines hemizygous for the modified locus.

Histology. After individual 12.5-dpc embryos had been bled into PBS, whole blood cells were gently pelletted and then fixed in 0.25% glutaraldehyde before staining with 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) at 37 °C for 24 h as described previously[20].

Nuclear run-on and endonuclease protection assays. To prepare membranes for nuclear run-on analysis, probe fragments were excised from plasmid vector by digestion with the appropriate restriction enzymes and purified by agarose-gel electrophoresis. Probe fragments used were an *HinfI–HinfI* fragment containing +17 to +956 relative to the transcription start site of the mouse α-globin gene; an *XbaI–PstI* fragment containing +150 to +969 of the mouse ζ-globin gene; and a *PstI–EcoRV* fragment of pSDK-*lacZ*pA[16], which includes +39 to +1164 relative to the start site of the *lacZ* transcript produced by αPElox. Purified fragment DNA (0.2 μg of each fragment) was then electrophoresed on 1.2% agarose gel and transferred to a nitrocellulose membrane by Southern blotting. Nuclei were purified as described[2] from 1.5×10^7 primitive erythroid cells collected from 12.5-dpc transgenic embryos. Run-on transcripts were synthesized from isolated nuclei and hybridized to membranes as described previously[21]. For endonuclease protection assays, nuclei prepared as described above were separated into four aliquots of 250 μl and digested with 0, 40, 80 or 160 U of *Eco*0109 I at 37 °C for 90 min. After proteinase-K digestion and phenol-chloroform extraction, purified DNA was digested with *EcoRI* and Southern analysis performed.

Acknowledgements

We thank P. Vassalli for the gift of the pCAGGS-Cre plasmid, H. Sutherland for the gift of the αPE plasmid, W. Magis for technical support and M. Groudine for support and helpful discussions. This work was supported by a grant from the National Health and Medical Research Council of Australia (to E.W.), an Australian Postgraduate Award (to D.G.), an NIH grant (to D.I.K.M.) and a grant from the Burroughs-Wellcome Fund to S.F. D.I.K.M. is a scholar of the Leukemia Society of America.

Received 6 August; accepted 10 October, 1997.

1. Linn, F., Heidmann, I., Saedler, H. & Meyer, P. Epigenetic changes in the expression of the maize A1 gene in *Petunia hybrida*: role of numbers of integrated gene copies and state of methylation. *Mol. Gen. Genet.* **222**, 329–336 (1990).
2. Scheid, O.M., Paszkowski, J. & Potrykus, I. Reversible inactivation of a transgene in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **228**, 104–112 (1991).
3. Assaad, F.F., Tucker, K.L. & Signer, E.R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* **22**, 1067–1085 (1993).
4. Dorer, D.R. & Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77**, 1–20 (1994).
5. Sabl, J.F. & Henikoff, S. Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in Drosophila. *Genetics* **142**, 447–458 (1996).
6. Davis, B.P. & MacDonald, R.J. Limited transcription of rat elastase I transgene repeats in transgenic mice. *Genes Dev.* **2**, 13–22 (1988).
7. Dorer, D.R. Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res.* **6**, 3–10 (1997).
8. Sharpe, J.A. *et al.* Analysis of the human α-globin gene cluster in transgenic mice. *Proc. Natl. Acad. Sci. USA* **90**, 11262–11266 (1993).
9. Higgs, D.R. *et al.* A major positive regulatory region located far upstream of the human α-globin gene locus. *Genes Dev.* **4**, 1588–1601 (1990).
10. Robertson, G. *et al.* Position-dependent variegation of globin transgene expression in mice. *Proc. Natl. Acad. Sci. USA* **92**, 5371–5375 (1995).
11. Hoess, R.H. & Abremski, K. Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. *Proc. Natl. Acad. Sci. USA* **81**, 1026–1029 (1984).
12. Lakso, M. *et al.* Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl. Acad. Sci. USA* **93**, 5860–5865 (1996).
13. Palmiter, R.D. & Brinster, R.L. Germ-line transformation of mice. *Annu. Rev. Genet.* **20**, 465–499 (1986).
14. Araki, K., Araki, M., Miyazaki, J.-I. & Vassalli, P. Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase. *Proc. Natl. Acad. Sci. USA* **92**, 160–164 (1995).
15. Robertson, G., Garrick, D., Wilson, M., Martin, D.I.K. & Whitelaw, E. Age-dependent silencing of globin transgenes in the mouse. *Nucleic Acids Res.* **24**, 1465–1471 (1996).
16. Sutherland, H.G.E., Martin, D.I.K. & Whitelaw, E. A globin enhancer acts by increasing the proportion of erythrocytes expressing a linked transgene. *Mol. Cell. Biol.* **17**, 1607–1614 (1997).
17. Meyer, P., Heidmann, I. & Niedenhof, I. Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. *Plant J.* **4**, 89–100 (1993).
18. Neuhuber, F., Park, Y.-D., Matzke, A.J.M. & Matzke, M.A. Susceptibility of transgene loci to homology-dependent gene silencing. *Mol. Gen. Genet.* **244**, 230–241 (1994).
19. Ye, F. & Signer, E.R. RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. *Proc. Natl. Acad. Sci. USA* **93**, 10881–10886 (1996).
20. de Carvalho, F. *et al.* Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *EMBO J.* **11**, 2595–2602 (1992).
21. Ingelbrecht, I., Van Houdt, H., Van Montagu, M. & Depicker, A. Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl. Acad. Sci. USA* **91**, 10502–10506 (1994).
22. Bresnick, E.H. & Felsenfeld, G. Dual promoter activation by the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA* **91**, 1314–1317 (1994).
23. Martin, D.I.K. & Whitelaw, E. The vagaries of variegating transgenes. *Bioessays* **18**, 919–923 (1996).
24. Grosveld, F., Van Assendelft, G.B., Greaves, D.R. & Kollias, G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).
25. Ryan, T.M. *et al.* A single erythroid specific DNase I superhypersensitive site activates high levels of human β-globin gene expression in transgenic mice. *Genes Dev.* **3**, 314–323 (1989).
26. Matzke, M.A., Priming, M., Trnovsky, J. & Matzke, A.J.M. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**, 643–649 (1989).
27. Rhounim, L., Rossignol, J.-L. & Faugeron, G. Epimutation of repeated genes in *Neurospora crassa*. *EMBO J.* **11**, 4451–4457 (1992).
28. Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett.* **146**, 148–152 (1982).
29. Garrick, D., Sutherland, H., Robertson, G. & Whitelaw, E. Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. *Nucleic Acids Res.* **24**, 4902–4909 (1996).
30. Pondel, M.D., Proudfoot, N.J., Whitelaw, C. & Whitelaw, E. The developmental regulation of the human ζ-globin gene in transgenic mice employing β-galactosidase as a reporter gene. *Nucleic Acids Res.* **20**, 5655–5660 (1992).
31. Whitelaw, E. & Proudfoot, N. α-Thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human α2 globin gene. *EMBO J.* **5**, 2915–2922 (1986).

TAB 8

Cell, Vol. 42, 519–526, September 1985, Copyright © 1985 by MIT 0092-8674/85/090519-08 $02.00/0

Negative Regulation of Viral Enhancers in Undifferentiated Embryonic Stem Cells

Cornelia M. Gorman,* Peter W. J. Rigby, and David P. Lane
Cancer Research Campaign
Eukaryotic Molecular Genetics Research Group
Department of Biochemistry
Imperial College of Science and Technology
London, SW7 2AZ, England

Summary

Many viral genomes, including those of SV40 and MuLV, are not efficiently expressed in undifferentiated embryonal carcinoma (EC) cells but are expressed in differentiated derivatives. This regulation appears to be at the level of transcription. We have used DNA-mediated gene transfer to analyze the function of several viral promoters in EC cells. We show that the SV40 early promoter works efficiently in an enhancer-independent fashion following transfection into undifferentiated cells. Strikingly, the promoter in the LTR of MSV does not function in such cells; but when upstream sequences, including the enhancer, are deleted expression ensues. Replacement of the SV40 enhancer by that of MSV results in inactivation of the SV40 early promoter in these cells. We propose that the undifferentiated cells contain a *trans*-acting regulatory factor (or factors) that reduces transcription by interacting with viral enhancers.

Introduction

Some viruses, for example Simian virus 40 (SV40), polyoma virus, and Moloney murine leukemia virus (MuLV), are unable to efficiently express their genomes in undifferentiated embryonal carcinoma (EC) cells (Linnenbach et al., 1980; Segal and Khoury, 1979; Stewart et al., 1982; Swartzendruber and Lehman, 1975; Teich et al., 1977). The induction of differentiation, which is accompanied by changes in the expression of cellular genes, results in the removal of the block to viral gene expression. It has been suggested that the lack of expression of the SV40 early proteins is due to post-transcriptional effects, at the level of the splicing of the early mRNAs (Segal et al., 1979), but the majority of the evidence indicates that the defect is in transcription from the viral promoters. The clearest evidence is available for polyoma. Mutants selected for the ability to grow on undifferentiated EC cells carry mutations in the region required for the efficient transcription of the early genes, and in a number of cases these mutations have been mapped to the enhancer (Fujimura and Linney, 1982; Katinka et al., 1981; Sekikawa and Levine, 1981).

Differentiation presumably requires alterations in the activities of proteins that regulate the expression of cellular genes. The simplest interpretation of the viral data is that differentiation induces a protein (or proteins) required for efficient utilization of the viral promoters or that it leads to the disappearance of a factor (or factors) that represses these promoters.

In contrast to papovaviruses and retroviruses, human adenoviruses replicate in undifferentiated EC cells (Imperiale et al., 1984). The E1a region of adenoviruses encodes proteins required for the subsequent expression of other adenovirus transcription units (Berk et al., 1979; Jones and Shenk, 1979). E1a products can also activate chromosomal genes (Kao and Nevins, 1983) and serve to increase transcription from a variety of promoters in transient expression assays (Gaynor et al., 1984; Green et al., 1983; Treisman et al., 1983). Mutations in the region encoding this *trans*-acting transcriptional activator prevent the growth of adenoviruses on differentiated cell lines because the other viral genes are not efficiently expressed. However, E1a mutants replicate in undifferentiated EC cells, suggesting that such cells contain an activity that can substitute for E1a (Imperiale et al., 1984).

We have used DNA-mediated gene transfer to analyze the regulation of papovavirus and retrovirus promoters in EC cells. Our results suggest strongly that the undifferentiated cells contain a regulatory factor (or factors) that operates on sequences within the enhancer region to reduce transcription from adjacent promoters.

Results

Infection of EC Cells by SV40 Virus

In order to be able to compare our results with previous work we infected undifferentiated F9 cells with SV40. Following high multiplicity infection we monitored expression of the viral early protein large T-antigen by immunocytochemical staining with a monoclonal antibody, PAb419, which recognizes an N-terminal epitope (Harlow et al., 1981). In agreement with earlier results, only a small number (0.01%) of morphologically distinct, and presumably differentiated, cells stained 72 hr after infection (Figure 1A). If, however, immediately after infection the cells were treated with retinoic acid to induce differentiation, T-antigen staining was not apparent 48 hr after infection, but 20%–50% of the nuclei stain 72 hr after infection (Figure 1B). Figure 1C shows a parallel culture of retinoic acid-treated, infected cells stained with the antikeratin monoclonal antibody LE61 (E. B. Lane, 1982) which specifically stains differentiated cells (E. B. Lane and B. L. M. Hogan, personal communication). Similar results were obtained with other EC cell lines, P19 and PCC4.

These data confirm that undifferentiated cells do not express the SV40 early region and that differentiation removes the block to expression.

Transfection of EC Cells with SV40 DNA

We transfected undifferentiated F9, P19, and PCC4 cells

* Present address: Department of Molecular Biology, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080

This material may be protected by Copyright law (Title 17 U.S. Code)

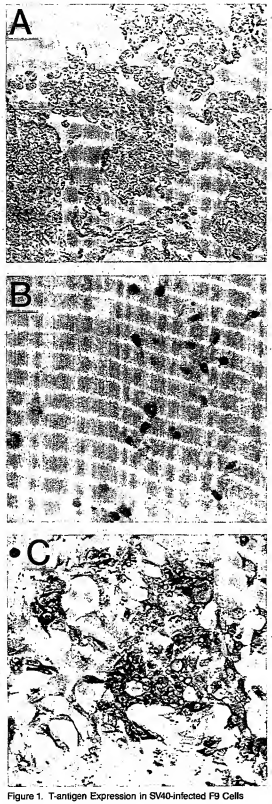


Figure 1. T-antigen Expression in SV40-infected F9 Cells

Three parallel cultures of undifferentiated F9 cells were infected with SV40 virus at a multiplicity of 20 pfu/cell. Immediately after infection two of the cultures (B and C) were induced to differentiate by the addition of retinoic acid while the remaining culture (A) was left untreated. Seventy-two hours after infection all three cultures were washed, fixed, and stained using the immunoperoxidase technique. (A) T-antigen staining of the undifferentiated cultures using PAb419 (10 μg/ml). (B) T-antigen staining of the differentiated cultures using PAb419 (10 μg/ml). (C) Keratin staining of the differentiated cultures using LE61 (10 μg/ml).

with pTSV3, a plasmid that contains the entire SV40 genome, using the calcium phosphate coprecipitation procedure. Forty hours after transfection we stained the cells with PAb419, with anti-SSEA-1 (Solter and Knowles, 1978), which detects a glycolipid found only on undifferentiated cells, with LE61, and with TROMA-1 (Kemler et al., 1981), another anticytokeratin antibody which stains only differentiated cells. As expected, all the cells stained with anti-SSEA-1, and only 0.1% stained with LE61 and TROMA-1, but to our surprise 30%–60% of the cells stained with PAb419.

In view of earlier reports that SV40 early RNAs are not properly spliced in undifferentiated EC cells, the use of a monoclonal antibody against an N-terminal epitope is not convincing proof of normal early region expression, as this antibody also recognizes small t-antigen which could be translated off an unspliced message. We therefore transfected undifferentiated F9 cells with pTSV3, and 40 hr later stained them with PAb419 (Figure 2A), with PAb402, which recognizes an epitope in the center of large T-antigen, and with PAb423 (Figure 2B), which recognizes a C-terminal epitope (Harlow et al., 1981). Staining was observed with all three antibodies. Figure 2C shows clear positive staining with PAb280, a monoclonal antibody specific for small t-antigen (Montano and D. P. Lane, 1984), proving that the cells are also expressing this protein. In parallel cultures stained with anti-SSEA-1 all the cells were positive whereas only 0.1% stained with TROMA-1 or LE61. When we transfected undifferentiated cells with uncloned SV40 DNA more than 50% of the nuclei stained for large T-antigen.

These data show that, when introduced into undifferentiated EC cells by calcium phosphate-mediated transfection, the SV40 early promoter functions efficiently. Thus there is no absolute block to its utilization. Although we have not directly analyzed the viral mRNAs in the transfected cells, the presence of an extreme C-terminal epitope of large T-antigen and of small t-antigen suggests strongly that the early mRNAs are properly spliced.

The level of large T-antigen expression depends on the amount of viral DNA transfected. If we used 1 μg of viral DNA plus 24 μg of carrier DNA, 5% of the undifferentiated cells stained; with 10 μg of viral DNA and 15 μg of carrier DNA, 20% of the cells stained, whereas with 25 μg of viral DNA 50% of the cells stained. This dependence on DNA concentration occurs only with viral DNA and is not seen in differentiated cells. Furthermore, if undifferentiated cells are transfected, immediately treated with retinoic acid, and then stained 72 hr later, this dose dependence for T-antigen staining is not observed. One interpretation of these results, discussed in more detail below, is that the large amounts of viral DNA introduced by transfection, as opposed to viral infection, titrate out a negative regulatory factor (or factors).

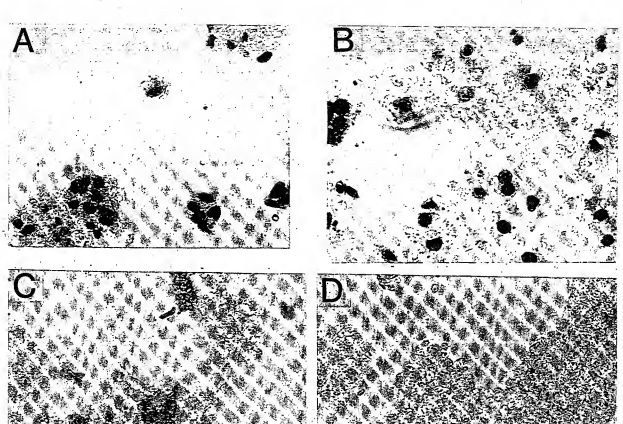


Figure 2. Expression of Large T- and Small t-antigens in Undifferentiated F9 Cells Transfected with SV40 DNA

Four replica cultures of undifferentiated F9 cells were transfected with 10 μg pTSV3 DNA. Forty hours later the plates were fixed and stained.
(A) Staining with PAb419, a monoclonal antibody that recognizes an epitope at or near the N terminus of Large T and small t.
(B) Staining with PAb423, a monoclonal antibody that recognizes an epitope at or near the carboxy terminus of large T and does not react with small t.
(C) Staining with PAb280, a monoclonal antibody that recognizes an epitope near the carboxy terminus of small t and does not react with large T.
(D) Staining with TROMA-1, a monoclonal antibody that recognizes an epitope expressed on keratin of differentiated but not of undifferentiated F9 cells.

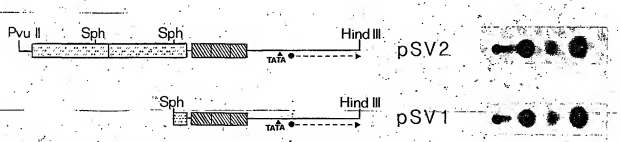


Figure 3. The SV40 Early Promoter Is Enhancer-Independent in Undifferentiated EC Cells

The structures of pSV2cat and pSV1cat are diagrammed. The former contains both the 72 bp repeat enhancer sequences (stippled boxes) and the GC-rich 21/22 bp repeats (hatched boxes), which are essential for promoter function. The latter lacks a functional enhancer. Also indicated are the locations of the TATA box and the start site for transcription.

The CAT assays shown are for extracts of undifferentiated F9 cells transfected with the two constructs, pSV2cat and pSV1cat. In this experiment 100 mm diameter dishes were transfected with 1 μg plasmid DNA plus 9 μg carrier DNA. One hundred microliter extracts were prepared and half of each was assayed for 1.5 hr at 37°C.

The amount of acetylated chloramphenicol was determined by cutting out the TLC spots and counting them in a scintillation counter. In each experiment pRSVcat was used as a reference plasmid. The amount of acetylated chloramphenicol produced in extracts of cells transfected with pRSVcat was taken as 100 and all other results are expressed relative to this. The numbers given in each figure legend are the average of the stated number of repetitions. Relative CAT activity [pSRVcat = 100]: pSV2, 20; pSV1, 19. This experiment was repeated ten times.

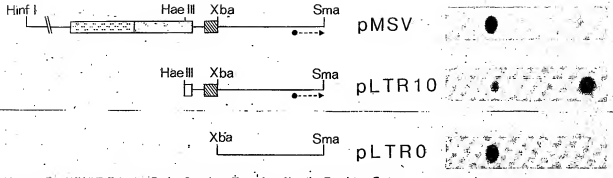


Figure 4. The MSV LTR Enhancer Region Contains a Target for a Negative Regulatory Factor

The structures of pMSVcat, pLTR10cat, and pLTR0cat are diagrammed. The construct pMSVcat contains the entire 560 bp LTR of MSV; indicated are the 73 bp (stippled box) and 72 bp (open boxes) enhancer sequences plus the GC-rich sequences (hatched boxes) next to the Xba I site. Plasmid pLTR10 lacks all of the 73 bp enhancer and most of the 72 bp enhancer, but retains the GC-rich region. Plasmid pLTR0 also lacks the GC-rich region and is transcriptionally inactive.

The CAT assays are on extracts of transfected, undifferentiated F9 cells. Sixty millimeter diameter dishes were transfected with 0.5 μg plasmid DNA plus 4.5 μg carrier DNA. One hundred microliter extracts were prepared of which 50 μl were assayed for 1 hr at 37°C. Relative CAT activity (pRSVcat = 100): pMSV, 1.5; pLTR10, 15; pLTR0, 0. This experiment was repeated six times.

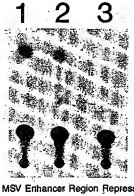


Figure 5. The MSV Enhancer Region Represses Transcription from the SV40 Early Promoter in Undifferentiated F9 Cells.

Undifferentiated F9 cells were transfected and CAT assays were performed as described in the legend to Figure 4. Lane 1, pSV2cat; Lane 2, pSVtcat; Lane 3, pSrM2cat. The arrow indicates a contaminant in the ^{14}C-chloramphenicol. Relative CAT activity (pSRVcat = 100): lane 1, 22; lane 2, 19; lane 3, 1.5. This experiment was repeated four times.

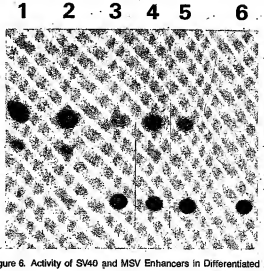


Figure 6. Activity of SV40 and MSV Enhancers in Differentiated F9 Cells

F9 cells were plated as a single cell suspension onto 60 mm diameter dishes. The cells were kept in medium containing 10^{-6} M retinoic acid for 4 days, being re-fed after 2 days. They were then fed again with retinoic acid-containing medium and transfected and assayed as described in the legend to Figure 4. Lane 1, pRSVcat; Lane 2, pSV2cat; Lane 3, pSV1cat; Lane 4, pMSVcat; Lane 5, pLTR10cat; Lane 6, pLTR0cat. Relative CAT activity (pRSVcat = 100): lane 1, 100; lane 2, 93; lane 3, 3; lane 4, 63; lane 5, 12; lane 6, 0. This experiment was repeated five times.

Transfection Does Not Induce Differentiation

Calcium can induce the differentiation of some cell types, for example keratinocytes (Hennings et al., 1983). We have therefore examined the effect of the calcium phosphate treatment involved in transfection on EC cell differentiation. Undifferentiated EC cells were treated with calcium phosphate precipitates lacking DNA for 12 hr, 18 hr, and 24 hr; the precipitate was then removed and the cells immediately stained. At all time points all the cells stained with anti-SSEA-1 and only 0.1% with LE61. After 24 hr exposure 0.1%–0.5% of F9 cells stained with TROMA-1.

We also investigated whether calcium phosphate treatment renders undifferentiated cells susceptible to infection by SV40 virus. Pretreatment with calcium phosphate caused no increase in the number of cells expressing large T-antigen. Treatment with calcium phosphate for 12 hr following virus infection led to the expression of large T-antigen in only 0.5% of the cells. Seventy-two hours of treatment with retinoic acid were required before large numbers (20%–50%) of cells stained for large T-antigen, even in the presence of calcium.

We therefore conclude that the expression of large T-antigen in transfected cultures is not due to induction of differentiation by the transfection procedure.

Promoter Utilization in Undifferentiated Cells

Undifferentiated cells were transfected with plasmids containing the easily assayable gene for chloramphenicol acetyl transferase (CAT) (Gorman et al., 1982b) driven by

Figure 7. Cotransfection of Undifferentiated F9 Cells

Transfections and CAT assays were performed as described in the legend to Figure 4. Lane 1, 2.5 μg pMSVcat plus 2.5 μg pRSVneo; Lane 2, 4 μg pMSVcat plus 1 μg pRSVneo; Lane 3, 5 μg pMSVcat. Relative CAT activity (pRSVcat = 100): lane 1, 75; lane 2, 40; lane 3, 1. This experiment was repeated three times.

a variety of viral promoters. The plasmid pRSVcat, which contains the promoter from the LTR of Rous sarcoma virus, directs the synthesis of high levels of CAT in all three cell lines studied—F9, P19, and PCC4—and was used as a standard in all experiments.

We investigated the role of a variety of viral transcriptional control elements in EC cells, using CAT assays to measure gene expression. In all cases we alter only upstream control sequences, not the transcribed region, and we therefore consider it extremely unlikely that any of the observed effects are due to changes in transcription termination, posttranscriptional processing, or RNA stability. In CV1 monkey cells deletion of the enhancer reduces expression from the SV40 early promoter 10- to 50-fold; in mouse fibroblasts the decrease is 10-fold. However, Figure 3 shows that there is no significant difference in the level of CAT expression in undifferentiated EC cells between pSV2cat, which contains the SV40 enhancer, and pSV1cat, which lacks it. We have therefore analyzed the activity of other viral enhancers in these cells. MuLV does not replicate in undifferentiated EC cells and Linney et al. (1984) have shown that the MuLV LTR does not function as a promoter in CAT transfection assays. We have analyzed the activity of the LTR of murine sarcoma virus (MSV), which was derived from MuLV, and have used three constructs: pMSVcat, which contains the entire LTR; pLTR10cat, which lacks sequences upstream of the promoter including 90% of the enhancer region; and pLTR0cat, which lacks all of the enhancer region plus a GC-rich region close to the cap site (see Figure 4). The plasmid pLTR0cat is completely inactive in fibroblasts. Activity cannot be restored by adding the SV40 enhancer, suggesting that the GC-rich region is required for promoter activity (Laimins et al., 1984).

Figure 4 shows the results of transfecting each of these plasmids into undifferentiated F9 cells. As expected, pLTR0cat is inactive. In agreement with the data of Linney et al. (1984) on the MuLV LTR, pMSVcat gives no, or very low, CAT activity; but surprisingly pLTR10cat, which lacks the enhancer region, directs the synthesis of appreciable levels of CAT. These data suggest that the MSV LTR does not function in these cells because they contain a negative regulatory factor (or factors) that acts on the enhancer region. To test this idea we used the plasmid pSrM2cat (Laimins et al., 1982), which contains the SV40 cap site and TATA box, but in which the 72 base pair repeats of the SV40 enhancer have been replaced by the 72/73 base pair repeat enhancer from the MSV LTR. We have shown above that the SV40 early promoter functions without its homologous enhancer in pSV1cat. The plasmid pSrM2cat is equivalent to the insertion of the MSV enhancer into pSV1cat. If the MSV enhancer region is the target for a negative regulatory factor (or factors), pSrM2cat should not function in undifferentiated EC cells. Figure 5 shows that this is the case. Plasmid pSrM2cat does function efficiently in mouse L cells, as would be expected from the known activity of the MSV enhancer in these cells (Laimins et al., 1982).

Promoter Utilization in Differentiated Cells

We used two procedures for transfection into differentiated F9 cells. Cells were either treated with retinoic acid for four days prior to transfection or transfected and then treated with retinoic acid for various times. Using the first protocol pSV2cat was 30 times more efficient than pSV1cat (Figure 6), indicating that differentiation restores the requirement for the SV40 enhancer. Similarly, pMSVcat gave higher levels of expression than pLTR10cat (Figure 6) showing that the negative regulation no longer occurs. When we transfected before inducing differentiation we observed substantial expression from the MSV LTR 48 hr after adding retinoic acid. By this time pSV2cat was being expressed more efficiently than pSV1cat.

Cotransfection Experiments

We had hoped to cotransfect marker plasmids with our experimental plasmids in order to standardize our experiments. This would be particularly important in comparisons between undifferentiated and differentiated cells, as the latter take up DNA much more efficiently. We therefore constructed pRSVβgal, in which transcription of the E. coli *lacZ* gene is driven by the RSV LTR promoter. However, our initial results were unsatisfactory; the amount of β-galactosidase synthesized was greatly affected by the promoter present on the cotransfected cat plasmid. For example, very high levels of β-galactosidase were synthesized in cells cotransfected with pRSVβgal and pLTR0cat or pSV1cat, whereas about half as much enzyme was present in cells co-transfected with pRSVβgal and pRSVcat.

We observed a more striking result when we assayed CAT activity in cells cotransfected with pMSVcat and constructs containing the RSV LTR. The plasmid pMSVcat is essentially inactive in undifferentiated F9 cells (see Figure 4), but Figure 7 shows that cotransfection of such cells with pMSVcat and pRSVneo, which contains the Tn5 aminoglycoside phosphotransferase gene under the control of the RSV LTR (Gorman et al., 1983), leads to high levels of CAT activity. CAT expression increases as the amount of pRSVneo DNA is increased. If pLTR0cat or pSV0cat is cotransfected with pRSVneo there is no effect of the cotransfected plasmid, suggesting that the high ex-

pression observed with pMSVcat does not result from recombination of the RSV LTR into the CAT plasmid. Taken together with the data reported above, these results indicate that the RSV LTR may compete for the negative regulatory factor (or factors), thus releasing the inhibition of the MSV LTR normally seen in undifferentiated F9 cells. Furthermore, preliminary experiments show that cotransfection of large amounts of pMSVneo with small amounts of pMSVcat appears to lead to significant CAT expression, again suggesting that the capacity of the cell to repress the MSV LTR is saturable.

Discussion

The data reported here confirm previous reports that the early transcription unit of SV40 is not expressed following viral infection of undifferentiated EC cells and that differentiation relieves the block to expression. However, introduction of the viral genome into undifferentiated cells by calcium phosphate-mediated transfection results in efficient expression. We have shown that the transfection procedure does not induce differentiation. The detection in transfected, undifferentiated cells of N- and C-terminal epitopes of large T-antigen and of small t-antigen makes it very unlikely that the viral transcripts are improperly spliced.

A major difference between viral infection and transfection is the number of copies of the viral genome introduced into each cell, which is much higher in the latter case. If the lack of expression following viral infection is due to a negative regulatory factor (or factors) analogous to a prokaryotic repressor, then one might expect that the introduction of large numbers of copies of the viral genome by transfection would titrate out the factor (or factors) and allow expression. This explanation is supported by our dose–response data, which show that increasing the proportion of SV40 DNA in a constant amount of total DNA transfected leads to an increase in the percentage of undifferentiated cells expressing large T-antigen. This effect is not seen in differentiated cells.

One problem with studies of gene expression using transient expression assays results from the variations in transfection efficiency between cell types and between experiments with a single cell type. Each of the experiments reported here has been done many times and we consistently see the same relative efficiencies of different control elements within the same cell type. We did, however, seek to control our experiments by cotransfecting pRSVβgal so that we could measure the levels of β-galactosidase expressed from the strong promoter in the RSV LTR and thus determine transfection efficiency. We chose the RSV LTR because our previous work had shown this promoter to be very strong in all cell types tested and we thus expected it to be neutral as far as the differentiation state of EC cells is concerned. However, pRSVβgal and pRSVneo clearly interact with the cotransfected plasmid, illustrating again the potential for competition for transcription factors in experiments of this type (Schöler and Gruss, 1984). Of particular interest in the present context is the observation that cotransfection with pRSVneo allows expression of pMSVcat in undifferentiated cells. The model proposed below provides a ready explanation for these data, namely that the RSV LTR competes for the negative regulatory factor (or factors) and thus relieves repression of the MSV LTR. Similar results have been obtained from experiments in which negative regulation of the SV40 enhancer by adenovirus E1a products could be titrated out by additional regulatory sequences (Borelli et al., 1984).

In all cell types previously tested efficient expression of the SV40 early region requires the presence of the viral enhancer or a heterologous enhancer element. However, our results show clearly that in undifferentiated EC cells, pSV1cat, which lacks the enhancer but retains the early promoter, is expressed as efficiently as the enhancer-containing construct pSV2cat. If there is a negative regulatory factor (or factors) in undifferentiated cells and it acts on sequences within the enhancer region, then pSV1cat would not be subject to repression. However, one still has to explain why the expression of pSV1cat is enhancer-independent. Perhaps a *trans*-acting factor present in undifferentiated EC cells can act to render the SV40 early promoter enhancer-independent. There is precedent for this idea in the behavior of 293 cells, which express the *trans*-acting E1a factor of adenovirus type 5, and in which a number of promoters do not require enhancers although they manifest a strict requirement for such elements in other cell types (Treisman et al., 1983).

These tentative conclusions from the SV40 experiments are strongly supported and extended by our studies on the MSV LTR. This viral control element is inactive in undifferentiated EC cells following both viral infection and DNA transfection. However, the deletion in pLTR10cat, which removes upstream sequences including the enhancer, results in expression following transfection into undifferentiated cells. This observation supports the hypothesis that undifferentiated cells contain a negative regulatory factor (or factors) which acts on sequences within the enhancer region, and that expression in undifferentiated cells is enhancer-independent. If the MSV enhancer region is inserted upstream of the SV40 early promoter in pSV1cat, expression in undifferentiated EC cells is markedly reduced, providing further evidence that the LTR enhancer region is the target for a cellular factor (or factors) preventing expression. Clearly the observed effect of the negative regulator (or regulators) will depend on its relative affinity for different enhancer sequences, on the intracellular copy number of the target sequence, and on the effect of binding on utilization of the adjacent promoter. Thus we hypothesize that at the high copy numbers achieved by transfection, the SV40 enhancer readily titrates all of the regulator, whereas the MSV enhancer does not. However, at the low copy numbers achieved by viral infection or by the transfection of small amounts of DNA, sufficient regulator is present to prevent expression of the SV40 early region. The competition experiments imply efficient binding of the factor by the RSV LTR; however, we must suppose that this binding is without significant effect on the RSV promoter.

Finally, close analysis of much of the data compiled for

polyoma virus expression supports our hypothesis. There are two enhancer regions in polyoma; one acts preferentially in fibroblasts, the other acts preferentially in embryonic cells (Herbomel et al., 1984). Deletion of the fibroblast enhancer can allow growth in F9 cells whereas mutations in the other region, which contains a GC-rich sequence possibly important for polyoma promoter function, can also allow expression in F9 cells. In the experiments in Linney et al. (1984) it is only this GC-rich region which renders a construct nearly identical to pLTR0cat competent for expression. The construct pLTR0cat itself has been shown to be completely transcriptionally inactive (Laimins et al., 1984).

Mutants of human adenovirus defective in E1a function do not replicate in differentiated cells because the other viral transcription units cannot be efficiently expressed in the absence of the *trans*-acting transcriptional activator encoded by the E1a region (Berk et al., 1979; Jones and Shenk, 1979). The ability of such mutants to replicate in undifferentiated EC cells suggests that these cells contain a factor that can substitute for E1a (Imperiale et al., 1984). It has recently been demonstrated that E1a proteins can also act as repressors of transcription, and that in the case of SV40 this repression is mediated by sequences within the enhancer region (Borelli et al., 1984; Velcich and Ziff, 1985).

We therefore conclude that undifferentiated EC cells contain a negative regulatory factor (or factors) that acts on viral enhancers to prevent expression of transcription units to which they are linked. It is this factor that prevents the expression of certain viral transcription units in undifferentiated cells; it must be destroyed, inactivated, or decrease in abundance during differentiation. It is possible that this factor either is, or is closely related to, the factor providing E1a function in undifferentiated cells. Presumably this factor plays a role in the maintenance of the undifferentiated state. It will be of great interest to characterize it further and to investigate its role in the control of cellular gene expression.

Experimental Procedures

Recombinant Plasmids

Plasmid pTSV3 contains the entire SV40 genome cloned as an Eco RI linear into the Eco RI site of pAT153 (Clayton et al., 1982). Plasmid pRSVcat was derived by the insertion of the LTR of Rous sarcoma virus into the standard cat vector pSV0cat; it has previously been shown that transcription of the cat gene in this plasmid originates at the correct start site within the RSV LTR (Gorman et al., 1982a). In pSV2cat, transcription of the cat gene is driven by the SV40 early promoter; correct transcription initiation has previously been demonstrated (Gorman et al., 1982a). Plasmid pSV1cat is derived from pSV2cat by the removal of the enhancer region (Gorman et al., 1982b). The construct pMSVcat is pMS–LTR2 of Laimins et al. (1984); transcription of the cat gene is driven by the Sma I–Hinf I fragment of the LTR, which contains the R and U3 regions. Plasmid pLTR10cat is pMS–LTR10 of Laimins et al. (1984); the enhancer region has been deleted. This construct retains 15 bp of the enhancer region and the GC-rich upstream sequences (see Figure 4). Plasmid pLTR0cat is pMS–LTR0 of Laimins et al. (1984); in this construct the whole of the enhancer and the GC-rich upstream sequences have been deleted (see Figure 4). The plasmid pRSVβgal was constructed from pCH110 (Hall et al., 1983) and pRSVcat. The Hind III–Bam HI fragment containing the *lacZ* gene and polyadenylation site from pCH110 was ligated to the Hind III–Bam HI fragment of pRSVcat, which contains the RSV LTR. In pSrM2cat transcription of the cat gene is driven by the SV40 early promoter, but the SV40 enhancer has been replaced by the enhancer element from the MSV LTR (Laimins et al., 1982). All recombinant plasmid DNAs were prepared according to Gorman et al. (1982b). DNAs to be used in transfection experiments were purified by two successive CsCl–ethidium bromide equilibrium density centrifugations, as we have found that this improves the reproducibility of the experiments.

Cell Culture, DNA-Mediated Gene Transfer, and Virus Infection

All cells were grown in Dulbecco's modification of Eagle's medium supplemented with 10% (v/v) fetal calf serum and antibiotics. Great care was taken to set the CO_2 concentration in the incubators such that the pH of the media was equilibrated at 7.3. The P19 line of EC cells was subcultured 1:10 every other day; the F9 and PCC4 lines were subcultured 1:10 every third day. Because of the extreme tendency of the F9 cells to clump they were subcultured with the aid of a 10 ml syringe and a 19 gauge needle. EC lines were maintained for 15 passages then discarded and new cells were recovered from storage.

We found it especially important to plate EC cells as single cell suspensions just prior to transfection. We modified the basic calcium phosphate coprecipitation technique to give optimal expression of various cat vectors, assaying expression either by CAT assays (Gorman et al., 1982b) or by immunocytochemical staining with anti-CAT monoclonal antibodies (C. M. Gorman and D. P. Lane, unpublished data). The best results were obtained by subculturing the EC cells 1:10 in the morning, allowing them to attach to the dish and then, about 5 hr after plating, adding the calcium phosphate–DNA coprecipitate which was left on the cells for 12–16 hr. The cells were then washed and fed with fresh medium. Cells were harvested, or fixed and stained, 24–36 hr later. The carrier DNA used in transfections was either pSV0cat (Gorman et al., 1982b) or pBR322.

To induce differentiation of F9 cells (Strickland and Mahdavi, 1978) they were subcultured so that a single cell suspension was obtained and retinoic acid was added to the medium at a final concentration of 10^{-7} M. The cells were re-fed every 48 hr with fresh medium containing retinoic acid. Differentiation was monitored by immunocytochemical staining.

Cells were infected with the wt830 strain of SV40 at an moi of 20 pfu/cell. The expression of large T-antigen was monitored by immunocytochemical staining at various times after infection. In control infections of NIH3T3 cells, 50% of the nuclei stained at 36 hr after infection.

Assays for CAT Expression

CAT enzyme assays were performed as described by Gorman et al. (1982b), except that only 1 μCi of ^{14}C-chloramphenicol (Amersham International) was used per assay. Some batches of ^{14}C-chloramphenicol contained a contaminant, which did not interfere with the assays; this is readily identified by running unmodified chloramphenicol on each TLC plate. For each assay we transfected a 60 mm dish with 5 μg DNA. Thirty-six hours after transfection the cells were harvested and lysed by freeze–thawing in 100 μl 0.25 M Tris–HCl (pH 7.8). Fifty microliters of the extract were added to 1 μCi ^{14}C-chloramphenicol, 20 μl 4 mM acetyl–CoA and 100 μl 0.25 M Tris–HCl (pH 7.8). Reactions were incubated at 37°C for 1 hr and stopped by the addition of 1 ml ethyl acetate. Quantitation was performed as described by Gorman et al. (1982b).

Immunocytochemical Staining

The following monoclonal antibodies were used. Anti-SSEA-1 (Solter and Knowles, 1978), recognizes a glycolipid determinant found on undifferentiated, but not on differentiated, EC cells. TROMA-1 (Kemler et al., 1979) recognizes a cytokeratin expressed only in differentiated cells, as does LE61 (E. B. Lane, 1982; E. B. Lane and B. L. M. Hogan, personal communication). PAb419, PAb402, and PAb423 recognize mapped determinants on SV40 large T-antigen (Harlow et al., 1981). PAb280 (Montano and D. P. Lane, 1984) recognizes a determinant unique to SV40 small t-antigen.

Cells were fixed using freshly made acetone:methanol (1:1) for 2 min. EC cells were fixed directly without washing. Antibodies were diluted in sterile PBS containing 10% (v/v) fetal calf serum. Hybridoma cell culture supernatants were used at a concentration of 5–10 μg/ml of antibody. Fixed cells were incubated with the first antibody for 3 hr

and then quickly washed five times in PBS. The second antibody, anti-mouse immunoglobulin conjugated with peroxidase (DAKO), was used at a dilution of 1:100. Incubation was for 1 hr followed again by five quick washes with PBS. O-dianisidine (made as a saturated ethanolic solution) was used diluted 1:100 in PBS with 0.01% H_2O_2 as the peroxidase substrate. Following a 30 min incubation in substrate, the cells were washed several times in distilled water.

Acknowledgments

We are grateful to Lou Laimins for plasmids, and Brigid Hogan and Keith Willison for EC cells. Anti-SSEA-1 was provided by Davor Solter, TROMA-1 by Rolf Kemler and Phillipe Brulet, and LE61 by Birgit Lane. We thank Barbara Skene for many helpful discussions and suggestions. CMG was supported by a NATO Post-Doctoral Fellowship and by a Long-Term Fellowship from the European Molecular Biology Organisation; P.W.J.R. holds a Cancer Development Award from the Cancer Research Campaign, which also paid for this work.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 3, 1985

References

Berk, A. J., Lee, F., Harrison, T., Williams, J., and Sharp, P. A. (1979). Pre-early adenovirus 5 gene product regulates synthesis of early viral messenger RNAs. Cell *17*, 935–944.

Borrelli, E., Hen, R., and Chambon, P. (1984). Adenovirus-2 E1a products repress enhancer stimulation of transcription. Nature *312*, 608–612.

Clayton, G. E., Lovett, M., and Rigby, P. W. J. (1982). Functional analysis of a Simian virus 40 super T antigen. J. Virol. *44*, 974–982.

Fujimura, F. K., and Linney, E. (1982). Polyoma mutants that productively infect F9 embryonal carcinoma cells do not rescue wild type polyoma in F9 cells. Proc. Natl. Acad. Sci. USA *79*, 1479–1483.

Gaynor, R., Hillman, D., and Berk, A. (1984). Adenovirus early region 1A protein activates transcription of a nonviral gene introduced into mammalian cells by infection or transfection. Proc. Natl. Acad. Sci. USA *81*, 1193–1197.

Gorman, C., Merlino, G., Willingham, M., Pastan, I., and Howard, B. (1982a). Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA mediated transfection. Proc. Natl. Acad. Sci. USA *79*, 6777–6781.

Gorman, C. M., Moffat, L. F., and Howard, B. H. (1982b). Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell Biol. *2*, 1044–1051.

Gorman, C. M., Padmanabhan, R., and Howard, B. H. (1983). High efficiency DNA-mediated transfection of mammalian cells. Science *221*, 551–553.

Green, M. R., Treisman, R., and Maniatis, T. (1983). Transcriptional activation of cloned human β-globin genes by viral immediate-early gene products. Cell *35*, 137–148.

Hall, C. V., Jacob, P. E., Ringold, G. M., and Lee, F. (1983). Expression and regulation of *Escherichia coli lacZ* gene fusions in mammalian cells. J. Mol. Appl. Genet. *2*, 101–109.

Harlow, E., Crawford, L. V., Pim, D. C., and Williamson, N. M. (1981). Monoclonal antibodies specific for Simian virus 40 tumor antigens. J. Virol. *39*, 861–869.

Hennings, H., Holbrook, K., and Yuspa, S. (1983). Factors influencing calcium induced terminal differentiation in cultured mouse epidermal cells. J. Cell Physiol. *116*, 265–281.

Herbomel, P., Bourchot, B., and Yaniv, M. (1984). Two distinct enhancers with different cell specificities coexist in the regulatory region of polyoma. Cell *39*, 653–662.

Imperiale, M. J., Kao, H. T., Feldman, L., Nevins, J., and Strickland, S. (1984). Common control of the heat shock gene and early adenovirus genes: evidence for a cellular E1a-like activity. Mol. Cell Biol. *4*, 867–874.

Jones, N., and Shenk, T. (1979). An adenovirus type 5 early gene function regulates expression of other early viral genes. Proc. Natl. Acad. Sci. USA *76*, 3665–3669.

Kao, H. T., and Nevins, J. R. (1983). Transcriptional activation and subsequent control of the human heat shock gene during adenovirus infection. Mol. Cell. Biol. *3*, 2058–2065.

Katinka, M., Vasseur, M., Montreau, N., Yaniv, M., and Blangy, P. (1981). Polyoma DNA sequences involved in control of viral gene expression in murine embryonal carcinoma cells. Nature *290*, 720–722.

Kemler, R., Brulet, P., Schnebelin, M. T., Gaillard, J., and Jacob, F. (1981). Reactivity of monoclonal antibodies against intermediate filament proteins during embryonic development. J. Embryol. Exp. Morphol. *64*, 45–60.

Laimins, L. A., Khoury, G., Gorman, C., Howard, B., and Gruss, P. (1982). Host-specific activation of transcription by tandem repeats from Simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. USA *79*, 6453–6457.

Laimins, L. A., Gruss, P., Pozzatti, R., and Khoury, G. (1984). Characterisation of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. *49*, 183–189.

Lane, E. B. (1982). Monoclonal antibodies provide specific intramolecular markers for the study of epithelial tonafilament organisation. J. Cell Biol. *92*, 665–673.

Linnenbach, A., Huebner, K., and Croce, C. M. (1980). DNA transformed murine teratocarcinoma cells: regulation of expression of Simian virus 40 tumor antigens in stem versus differentiated cells. Proc. Natl. Acad. Sci. USA *77*, 4875–4879.

Linney, E., Davis, B., Overhauser, J., Chao, E., and Fan, H. (1984). Non-function of a Moloney murine leukaemia virus regulatory sequence in F9 embryonal carcinoma cells. Nature *308*, 470–472.

Montano, X., and Lane, D. P. (1984). Monoclonal antibody to Simian virus 40 small t. J. Virol *51*, 760–767.

Schöler, H. R., and Gruss, P. (1984). Specific interaction between enhancer-containing molecules and cellular components. Cell *36*, 403–411.

Segal, S., and Khoury, G. (1979). Differentiation as a requirement for Simian virus 40 gene expression in F9 embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA *76*, 5611–5615.

Segal, S., Levine, A. J., and Khoury, G. (1979). Evidence for nonspliced SV40 RNA in undifferentiated teratocarcinoma stem cells. Nature *280*, 335–338.

Sekikawa, K., and Levine, A. J. (1981). Isolation and characterisation of polyoma host-range mutants that replicate in nullipotential embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA *78*, 1100–1104.

Solter, D., and Knowles, B. (1978). Monoclonal antibody defining a stage specific mouse embryonic antigen SSEA-1. Proc. Natl. Acad. Sci. USA *75*, 5565–5569.

Stewart, C. L., Stuhlmann, H., Jahner, D., and Jaenisch, R. (1982). De novo methylation, expression and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA *79*, 4098–4102.

Strickland, S., and Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell *15*, 393–403.

Swartzendruber, D. E., and Lehman, J. M. (1975). Neoplastic differentiation: interaction of Simian virus 40 and polyoma virus with murine teratocarcinoma cells in vitro. J. Cell. Physiol. *85*, 179–188.

Teich, N. M., Weiss, R. A., Martin, G. R., and Lowy, D. R. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell *12*, 973–982.

Treisman, R., Green, M., and Maniatis, T. (1983). Cis- and trans-activation of globin gene transcription in transient assays. Proc. Natl. Acad. Sci. USA *80*, 7428–7432.

Velcich, A., and Ziff, E. (1985). Adenovirus E1a proteins repress transcription from the SV40 early promoter. Cell *40*, 705–716.

TAB 9

JOURNAL OF VIROLOGY, Feb. 1991, p. 904–912
0022-538X/91/020904-09$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 65, No. 2

Inactivation of the Moloney Murine Leukemia Virus Long Terminal Repeat in Murine Fibroblast Cell Lines Is Associated with Methylation and Dependent on Its Chromosomal Position

ROB C. HOEBEN, ALEXANDRA A. J. MIGCHIELSEN, ROBERT C. M. VAN DER JAGT, HANS VAN ORMONDT, AND ALEX J. VAN DER EB*

Laboratory for Molecular Carcinogenesis, Sylvius Laboratories, University of Leiden, Wassenaarseweg 72, P.O. Box 9503, 2300 RA Leiden, The Netherlands

Received 20 August 1990/Accepted 5 November 1990

The expression of a retroviral vector with the Moloney murine leukemia virus (Mo-MuLV) long terminal repeat (LTR) promoter after integration into the genome of murine fibroblast cell lines was monitored with the *Escherichia coli*-derived β-galactosidase (β-gal) gene as the reporter. Monoclonal cell lines derived after retroviral infection exhibited a marked heterogeneity in their expression of the reporter gene. We studied two monoclonal cell lines with a single unrearranged copy of the vector provirus integrated into their genome. The first, BB10, expressed the marker enzyme in only 8% of its cell population, whereas in the second, BB16, β-gal expression could be detected in over 98% of the cells. Treatment of BB10 with the DNA-demethylating agent 5-azacytidine raised the number of β-gal-positive cells to over 60%. Transfection experiments showed that the Mo-MuLV LTR promoter-enhancer is potentially fully functional in both the BB10 and BB16 cell lines. The inactivated provirus from BB10 cells was cloned and subsequently used to generate retrovirus stocks. The promoter-enhancer activity of its LTR after infection with these BB10-derived viruses showed a variation similar to that of the original virus stocks. Our data showed that (1) inactivation of the Mo-MuLV LTR is a frequent event in murine fibroblast cell lines, (2) inactivation is associated with de novo methylation of cytidine residues, (3) the frequency of inactivation of the provirus must be determined by its chromosomal position, (4) the process of methylation of sequences within the LTR is not necessarily the same as the transcription-repression mechanism that is operating in undifferentiated embryonal carcinoma cells.

Downloaded from jvi.asm.org by on September 14, 2009

Retroviral vectors offer several advantages for gene transfer experiments. First, the highly efficient integration process results in the insertion of a DNA copy of the virus vector with a predictable structure into the host cell genome. Second, they can be used to infect a wide variety of host cells and may carry, apart from the gene of interest, a dominant selectable marker. Third, high-titer helper-free stocks of recombinant retroviruses can be obtained with relative ease. Therefore, retroviral vectors are widely used for experiments aimed at the development of genetic therapies for human genetic disorders (11).

So far, applications have been seriously hampered by the fact that the retroviral regulatory elements are repressed in certain cell types. It has been amply documented (2, 9, 27, 28, 35, 36, 52, 57) that the Moloney murine leukemia virus (Mo-MuLV) long terminal repeat (LTR) is expressed inefficiently in undifferentiated embryonal carcinoma (EC) cell lines and in preimplantation embryos. The restriction of expression occurs at several levels but is caused at least partially at the level of transcription initiation. Within the LTR, the enhancer sequences are targets for repressing factors (10, 14, 20, 26, 52). It has been shown that de novo methylation of CpG doublets is associated with this suppression (6, 12, 31, 50; reviewed in reference 55). Methylated LTRs are inefficient as promoters in murine fibroblast cell lines, whereas unmethylated LTRs are fully functional (19, 46).

An analogous or identical suppression mechanism has been suggested to operate after retrovirus-mediated gene transfer into hematopoietic stem cells (18, 29, 53). As a consequence, vectors that are functional in mature hematopoietic cells often are not expressed in blood cells of animals transplanted with infected stem cells (29, 54, 58). As hematopoietic stem cells are considered to be a potential target tissue for gene therapy, much effort has been put into the improvement of vectors to overcome the suppression phenomenon. Modification of the enhancer within the LTR has been exploited to prevent the suppression (53). The introduction of additional promoters into the viral transcription unit was partially successful, but expression levels in blood cells in vivo still remain variable in most cases (3, 54, 58). However, recent experiments suggest that failure to reconstitute the bone marrow of the recipient or failure to infect the donor stem cells may have contributed to the variable expression levels observed in vivo, thereby possibly overemphasizing the involvement of the transcription-suppression phenomenon (4, 24, 32, 59).

In an effort to gain more insight in the suppression mechanism, we used the retroviral vector BAG (37) (Fig. 1A), which contains the *Escherichia coli* β-galactosidase gene (β-gal) as a reporter to monitor the expression of the Mo-MuLV LTR promoter. The activity of β-gal can be visualized in individual cells by a sensitive in situ assay. The BAG vector also carries the Tn*5*-derived neomycin phosphotransferase gene (Neo^r), which is driven by the simian virus 40 early promoter and is therefore independent of the Mo-MuLV LTR activity. Here we report on the stability of the Mo-MuLV LTR-driven expression in murine fibroblast cell lines. We demonstrate that there is a marked heterogeneity in β-gal expression within monoclonal BALB/c 3T3 cell lines containing a single integrated copy of the BAG

* Corresponding author.

vector provirus. We show that inactivation of the Mo-MuLV LTR is directly associated with de novo methylation of cytidine residues. In addition, we demonstrate that the frequency of inactivation depends on the chromosomal position of the provirus. These results help to provide insight into the mechanisms involved in the down-regulation of transcription of the Mo-MuLV LTR.

MATERIALS AND METHODS

Cell lines and virus preparation. All sera and tissue culture media were purchased from GIBCO Laboratories (Grand Island, N.Y.). All culture plastics were obtained from Greiner (Nürtingen, Federal Republic of Germany). The hypoxanthine phosphoribosyltransferase-negative ($HPRT^-$) BALB/c 3T3 cell line B77 and the EC cell line P19 have been described before (30, 34). The psi-2-BAG cell line (37) (Fig. 1A), which produces helper-free ecotropic BAG viruses, was obtained from the American Type Culture Collection (ATCC 9560). This vector contains β-gal from *E. coli* driven by the Mo-MuLV LTR, together with the neomycin phosphotransferase gene from transposon Tn5 that renders bacteria and eucaryotic cells resistant to kanamycin and G418, respectively. Both cell lines were grown in high-glucose (4.5 g/liter) Dulbecco modified Eagle medium supplemented with 8% fetal calf serum in a 5% CO_2 atmosphere at 37°C. To isolate virus stocks, near-confluent cultures were grown in fresh medium for 16 h, after which the medium was isolated and filtered through a 0.45-μm-pore-size filter (Gelman Sciences, Inc., Ann Arbor, Mich.) and stored at −80°C until further use. Virus infections and titrations were performed as described previously (21); titers of the stocks ranged from 1×10^5 to 3×10^5 Neo^r CFU/ml when tested on B77 cells.

Cell lines containing a single proviral copy of the BAG virus were generated by infection of subconfluent B77 cell cultures with a low multiplicity of infection (MOI < 0.001) in the presence of 8 μg of Polybrene (Sigma) per ml. Twenty-four hours after infection, the cultures were split 1 in 10 in medium containing 400 μg of G418 (GIBCO) per ml. Twelve days postinfection, individual G418-resistant colonies were isolated, expanded, and assayed for β-gal expression. Individual cell lines derived in this manner were designated BB1 to BB19. These cell lines were routinely grown in medium containing 400 μg of G418 per ml. The demethylating agent 5-azacytidine (azaC) (Boehringer GmbH, Mannheim, Federal Republic of Germany) was used at 4 μM concentration unless stated otherwise.

Southern blot analysis. High-molecular-weight cellular DNA was isolated, and 10-μg samples were digested for 16 h in 400 μl with 40 units of *Hin*dIII or *Sma*I (Pharmacia, Uppsala, Sweden). After 60 min, a 20-μl sample was removed from the reaction mixture and added to 0.5 μg of adenovirus type 2 DNA and incubated for a further 15 h. The restriction fragments of the latter incubation were resolved by electrophoresis on a 1% agarose gel. Complete digestion of the adenovirus DNA indicated complete digestion of the chromosomal DNA. The digested chromosomal DNA was size fractionated on a 1% agarose gel and transferred to a GeneScreen Plus membrane (Dupont, NEN Research Products) by standard procedures (40). Hybridization was performed to either a Neo^r-specific probe (the 1,321-bp *Hin*dIII-*Sma*I fragment of pRSVneo [13]) or a β-gal-specific probe (the 154-bp *Bam*HI-*Bgl*I fragment of plasmid pBAG, encoding the amino terminus of the β-gal protein). Probes were labeled to high specific activity with [α-^{32}P]dCTP by the random primer method (8).

RNA analysis. Cytoplasmic RNA was isolated from near-confluent cell cultures as described previously (43). Quantitation of specific transcripts was performed after transfer of serial dilutions of RNA onto nitrocellulose membrane via the slot-blot procedure (Schleicher & Schuell, Dassel, Federal Republic of Germany) and hybridization with specific probes. Apart from the probes mentioned above, a nearly full-length cDNA clone of human elongation factor 1 (5) was also used as a positive control.

β-Gal cytochemistry. To reveal the bacterial β-gal activity in situ, cells were fixed and stained with 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal; Research Organics Inc., Cleveland, Ohio) essentially as described by Sanes et al. (41). Cells were incubated in the staining solution for 14 to 20 h before enumeration of the fraction of β-gal-positive cells. With the use of a standard bright-field microscope, a minimum of 500 cells were scored in all cell lines tested.

Recombinant DNA techniques. All recombinant DNA techniques followed standard protocols (40). To clone the inactivated provirus from the BB10.c37 cells, genomic DNA was isolated and 10 μg was digested to completion with *Nhe*I. Subsequently, the digested DNA was recovered and circularized with T4 DNA ligase (Boehringer) in a volume of 1 ml. The ligated DNA (3 μg) was used for transformation into the transformation-competent *E. coli* MC1061 (40) and plated onto agar plates containing 50 μg of kanamycin per ml. A clone containing a plasmid (pBB10c37.1) with the desired structure was used for further study.

CAT plasmids and assay. A 10-μg sample of the plasmid to be tested was transfected into 2×10^5 cells by the calcium phosphate precipitation procedure (15) in a 10-cm tissue culture dish. Plasmid pMuLV-CAT was constructed by inserting the *Nhe*I-*Sma*I fragment from the plasmid pBAG (kindly provided by C. Cepko, Harvard Medical School, Boston, Mass.), containing almost the complete U3 region and part of the R region of the MuLV LTR (nucleotides 7846 to 8292; numbered according to reference 45) into plasmid pBL-CAT5 (48) (kindly supplied by R. Offringa). This plasmid was digested with *Bgl*II, and the sticky ends were filled in with DNA polymerase (Klenow fragment) and digested with *Xba*I prior to ligation with the BAG *Nhe*I-*Sma*I fragment. Plasmids pBL-CAT5 and pRSV-CAT (14) were included as negative and positive controls, respectively. Forty hours after transfection, the cells were lysed and 50 μg of the protein lysate (100 μg for P19 cells) was used for the chloramphenicol acetyltransferase (CAT) activity assays. These were performed as described previously (13). The incubation time of the assays was 60 min (120 min for P19 cells).

RESULTS

β-Gal is expressed heterogeneously within BAG virus-infected clones. We studied the Mo-MuLV LTR promoter activity in murine fibroblast cell lines infected with a recombinant retroviral vector by exploiting several features of the BAG vector (Fig. 1A). In a first series of experiments, the percentage of β-gal-expressing cells was determined in 135 independent BAG-infected $G418^r$ BALB/c 3T3 (B77) colonies. After staining, most of the colonies showed a mosaic phenotype with respect to β-gal expression (Fig. 1B). A large portion of the colonies contained considerable numbers of β-gal-negative cells. The results are summarized in Fig. 1C. A similar heterogeneity was observed in a number of other BAG-infected murine fibroblast cell lines, including Swiss 3T3 and PA317 (20a). In a control experiment, colonies

Downloaded from jvi.asm.org by on September 14, 2009

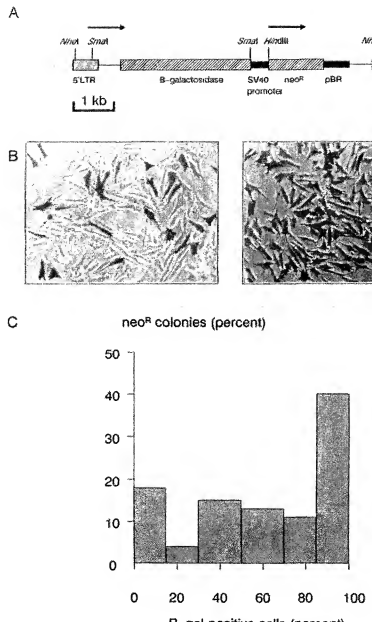


FIG. 1. (A) Schematic representation of the BAG vector used in this study. The 5' ends of the transcripts encoded by this vector are indicated by horizontal arrows. Both transcripts end at the polyadenylation site in the 3' LTR. Relevant restriction sites are indicated. A detailed description of this vector is given by Price et al. (37). pBR indicates the pBR322-derived origin of replication. (B) Photomicrographs of two BAG virus-infected and G418-resistant B77 clones. B77 cells were infected with BAG virus at low MOI and grown in medium containing 400 μg of G418 per ml for 12 days to select for infected cells. After selection, the cells were fixed and stained for the bacterial β-gal as described previously (41). Phase-contrast photomicrographs of two lines representing extreme phenotypes are shown (left, BB10; right, BB16; see text). (C) Histogram representing the fraction of β-gal-positive cells in 135 monoclonal BAG-infected B77 cell clones. B77 cells were infected at low MOI with BAG virus. Following selection with 400 μg of G418 per ml for 12 days, the resulting colonies were stained for the bacterial β-gal activity. The fraction of β-gal-positive cells was determined microscopically. On the ordinate, the number of colonies is shown as the percentage of the total number of colonies analyzed.

derived from the psi-2-BAG virus-producing cell line were assayed. In the latter experiment, all colonies consisted exclusively of β-gal-positive cells, excluding the possibility that the observed heterogeneity is due to improper fixation or staining.

To study the heterogeneity within the mosaic cell lines in more detail, we isolated 20 independent BAG-infected B77 cell lines. In these cell lines, the percentage of β-gal-expressing cells ranged between 8 and >98%. Two clones with the extreme phenotypes, BB10 (containing 8% β-gal-positive cells) and BB16 (>98% of the cells β-gal positive) were used for further study. Southern analysis established

Downloaded from jvi.asm.org by on September 14, 2009

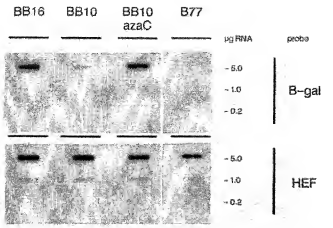


FIG. 2. Slot-blot quantitation of the β-gal-specific transcripts in B77, BB16, and BB10 cells and in BB10 cells treated with azaC. Cytoplasmic RNA was isolated from near-confluent cultures of B77, BB10, and BB16 cells and in addition from BB10 cells that had been treated with 4 μM azaC for 72 h. Serial dilutions of cytoplasmic RNA (5, 1, and 0.2 μg) were applied onto a nitrocellulose membrane with a slot-blot apparatus and hybridized with a β-gal-specific probe. As a control, a parallel filter was hybridized with a human elongation factor 1-specific (HEF) probe.

that both BB10 and B16 carried a single unrearranged copy of the vector provirus in their genomes (data not shown). A screening of the cytoplasmic RNA isolated from BB10, BB16, and the parental B77 cells for β-gal-specific sequences showed that the hybridization signal correlates with the relative abundance of the number of β-gal-positive cells (Fig. 2). The high frequency of inactivation suggests that the shutdown of β-gal expression is not the result of mutation of vector sequences but of some epigenetic mechanism.

Inactivation of LTR is reversed by addition of azaC. It is now well documented that in undifferentiated EC cells the block in transcription initiation is associated with the methylation of cytidine residues in CpG dinucleotides in the Mo-MuLV LTR (reviewed in reference 55). To study whether the Mo-MuLV LTR promoter could be reactivated by demethylation in the BB10 cells, we grew this cell line in various concentrations of the DNA-demethylating agent azaC for 3 days and determined the percentage of β-gal-positive cells. The fraction of β-gal-positive BB10 cells increased in a dose-dependent manner (Fig. 3A), whereas azaC did not change β-gal expression in B77 (0% β-gal positive) and in BB16 (>98% β-gal positive). Maximum stimulation in BB10 cells was achieved at 4 μM azaC (Fig. 3A); higher concentrations severely reduced the viability of the cells. These results indicate that the heterogeneity in the BB10 cell line is associated with de novo methylation of cytidine residues.

In a parallel experiment, a polyclonal cell culture of BAG virus-infected and G418-selected B77 cells containing 33% β-gal-positive cells was grown in 4 μM azaC for 72 h. This increased the number of β-gal-positive cells to 80%, indicating that the methylation-associated shutdown is a general phenomenon in BAG-infected B77 cells and not a peculiarity of the BB10 cell line. Thus, BB10 is a useful model for studying LTR inactivation in murine fibroblast cell lines.

Apart from the effect on the methylation of DNA, azaC also reduces DNA synthesis (25). To verify that the observed effect results from demethylation rather than reduction of DNA synthesis, we treated BB10 cells for 3 days with various concentrations of 1-β-D-arabinofuranosylcytosine-5′-triphosphate, a known inhibitor of DNA polymerases α and β (61). No alterations were seen in the fraction of β-gal-positive cells, even at concentrations which reduced [^{3}H]thymidine incorporation to less than 10% (data not shown).

To study whether the azaC-induced expression was stable in time, we grew BB10 cells with various concentrations of azaC for 72 h. After this treatment, the cell population was cultured in G418-containing culture medium. At regular intervals, subcultures were analyzed for β-gal expression. The results are shown in Fig. 3B. After the initial increase caused by azaC, the fraction of β-gal-expressing cells in the

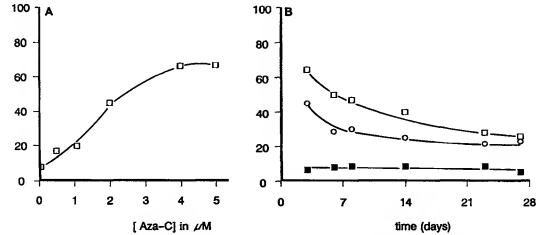


FIG. 3. (A) azaC-dependent increase in the relative number of β-gal-expressing cells. Confluent BB10 cultures were split 1 in 10 and subsequently grown in the presence of various concentrations of azaC for 72 h. After fixation, the cultures were stained for β-gal. The percentage of β-gal-positive cells was determined by microscopically scoring at least 500 cells. (B) Decrease in the fraction of β-gal-positive cells after azaC treatment of BB10 cells. Confluent BB10 cultures were split 1 in 10 and grown for 72 h in the presence of 4 μM (□) or 2 μM (○) azaC or in the absence of azaC (■) to induce β-gal expression. Subsequently, the cells were fed fresh medium without azaC, and at regular intervals subcultures were fixed and stained for β-gal activity.

Downloaded from jvi.asm.org by on September 14, 2009

culture gradually decreased. These results show that the vector provirus LTR again becomes inactivated after having been induced by azaC. Thus, inactivation of the LTR is independent of the integration process of the virus. A similar decrease was seen in parallel cultures grown without G418, excluding the possibility that the decrease was induced by the selection (data not shown). Experiments in which BB16 and B77 cells were grown in mixed cultures did not reveal significant differences in growth rates or survival after azaC treatment between β-gal-positive and -negative cells (data not shown).

Helper viruses are not involved in azaC-induced reactivation. The murine genome contains many copies of defective endogenous retroviruses. The majority of these are heavily methylated and thus transcriptionally inactive (reviewed in reference 7). There is evidence that in BALB/c mice, as the result of recombination between defective endogenous retroviruses, replication-competent viruses can be generated (7). Conceivably, the same could occur in the azaC-treated cell lines in which the BAG vector might be rescued from the initially β-gal-positive cells. Subsequent reinfection of the β-gal-negative cells might then contribute substantially to the increase of the fraction of β-gal-expressing cells. Several types of experiments were performed to test this hypothesis. First, no reverse transcriptase activity could be detected in the conditioned medium from BB10 and BB16 cells after azaC treatment of the cells. Second, culture medium from azaC-treated cells could not confer G418 resistance to B77 cells (<0.2 G418^r CFU/ml), nor could it induce β-gal expression after infection of B77 cells (<0.2 β-gal$^+$ induction units per ml). And third, no additional integration sites could be detected in a Southern analysis of *Hin*dIII-digested genomic DNA from nine subclones isolated after azaC treatment of BB10 cells (Fig. 4). These results clearly demonstrate that rescue of the BAG virus is not responsible for the increase of the fraction of β-gal-positive cells resulting from azaC treatment.

Sequences within the 5′ LTR of BB10 cells but not in that of BB16 cells are methylated. To determine whether sequences within the 5′ LTR of the BAG vector provirus are methylated, we exploited the fact that the LTR contains a site for the restriction endonuclease *Sma*I. This enzyme is inhibited when the inner cytidine residue of its recognition sequence, 5′-CCCGGG-3′, is methylated (38). *Hin*dIII-digested chromosomal DNA from BB10 (8% β-gal-positive cells), BB16 (>98% β-gal-positive cells), and B77 (parent) cells was digested for 16 h with excess *Sma*I. After transfer of the restriction fragments to a nylon membrane, the blot was hybridized with a β-gal-specific probe (Fig. 5). As expected, no signal could be detected in the lanes containing B77 DNA. In the BAG virus-infected cell lines BB10 and BB16, a single hybridizing fragment was detected in the *Hin*dIII-digested samples. Since the vector contains only a single *Hin*dIII site, the size of this fragment depends on the location of the nearest *Hin*dIII site in the bordering chromosomal DNA and thus differs between the BB10 and BB16 cell lines. In BB16, a single hybridizing fragment was detected in the *Hin*dIII-*Sma*I-digested sample, comigrating with the corresponding fragment of *Sma*I-digested pBAG DNA. This indicates that in BB16 neither *Sma*I site flanking the β-gal gene is methylated. In BB10 DNA, however, an additional fragment is seen in the *Hin*dIII-*Sma*I-digested DNA. Since the size of this fragment approximates the size of the corresponding *Hin*dIII fragment, we conclude that the *Sma*I site in the 5′ LTR is refractory to cleavage in a part of the DNA molecules. On the basis of the intensity of the signals,

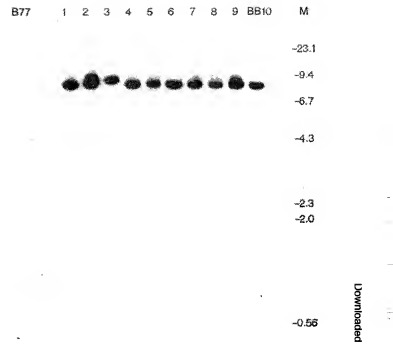


FIG. 4. Southern analysis of BB10 subclones isolated after azaC treatment. BB10 cells were treated with 4 μM azaC for 72 h. Following this treatment, subclones were isolated after limiting dilution of the cell population. From uninfected B77 cells, untreated BB10 cells, and the nine subclones (lanes 1 to 9), genomic DNA was isolated and 10-μg samples were digested with *Hin*dIII and subsequently fractionated on a 1% agarose gel. The Southern blot was hybridized with a Neo^r-specific probe. The positions of the size markers (M) are indicated (sizes in kilobases).

we estimate that in approximately 30% of the DNA molecules the *Sma*I site is protected by methylation. This is considerably lower than the fraction of β-gal-negative cells in the BB10 cell line, suggesting that the methylation of the *Sma*I site in the 5′ LTR is not the direct cause of the transcriptional shutdown but rather a delayed event.

Similar results have been reported in EC cell lines, in which the inactivation precedes the general methylation of proviruses by many days (12, 55). However, our observation that addition of azaC to the culture medium results in a rapid increase in expression of β-gal, as shown above, suggests that methylation of specific sequences within the LTR is somehow involved in the repression.

LTR enhancer-promoter sequences are potentially fully functional in B77 and derived cell lines. Our results suggest that, in contrast to the situation in EC cells, inactivation of the LTR can occur in cells in which the enhancer is potentially fully functional. To illustrate this difference, we determined the LTR promoter activity in both cell systems. The Mo-MuLV LTR was cloned in front of the CAT gene and introduced into B77, BB16, BB10, and P19 EC cells. The results are summarized in Table 1. As expected, the MuLV LTR was equally active in B77 and its BAG-infected derivatives BB10 and BB16, whereas it was repressed in P19 cells. Transfection of the pRSV-CAT plasmid induced a significant conversion of the substrate (see reference 14), showing that the lack of expression after transfection of

Downloaded from jvi.asm.org by on September 14, 2009

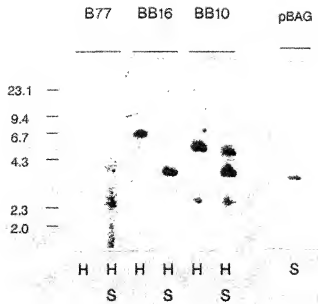


FIG. 5. Southern analysis of B77 clones infected with the BAG virus. Genomic DNA (10 μg) of BB10, BB16, and uninfected B77 cells was digested with *Hin*dIII (lanes H) or with *Hin*dIII and *Sma*I (lanes H and S) and fractionated on a 1% agarose gel. As a control, *Sma*I-digested plasmid pBAG DNA (10 pg) was used. The Southern blot was hybridized with a β-gal gene-specific probe. The positions of the size markers are indicated (sizes in kilobases).

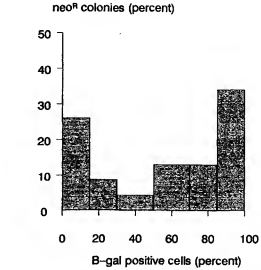


FIG. 6. Histogram representing the fraction of β-gal-positive cells in 23 monoclonal BB10c37 virus-infected B77 cell clones. B77 cells were infected at low MOI with BB10c37 virus. Following selection with 400 μg of G418 per ml for 12 days, the resulting colonies were stained for the bacterial β-gal activity. The fraction of β-gal-positive cells was determined microscopically. On the ordinate, the number of colonies is shown as the percentage of the total number of colonies analyzed.

pMuLV-CAT in P19 cells is not just caused by inefficient transfection. These observations indicate that the MuLV LTR promoter-enhancer is actively repressed in the undifferentiated EC cells, whereas in the B77 cells it is fully functional. Apparently, inactivation of the MuLV LTR after integration can occur in cell types in which the LTR is potentially fully functional.

Frequency of inactivation is chromosome-position dependent. As shown above, an epigenetic mechanism is responsible for the shutdown of expression in murine fibroblast cell lines. The heterogeneity within the cell lines is generated by temporal variations in the occurrence of the inactivating event in individual cells within a certain cell line. Apart from this heterogeneity within monoclonal cell lines, a large variation is seen between different cell lines with respect to the overall frequency of inactivation. This variation, illustrated in Fig. 1C, might be caused by the vicinity of regulatory elements in the flanking host cell genome, viz., a chromosome position effect. Alternatively, retroviruses are known to have a high mutation rate (47), and mutations might increase the susceptibility to methylation. To differentiate between these two hypotheses, we decided to clone the inactivated provirus from BB10 cells. To eliminate the β-gal-positive cells from the BB10 population, we derived several subclones which contained less than 1 β-gal-positive cell per 10^4. Genomic DNA of one of these subclones, BB10.c37, was used to clone the provirus. This is made possible by the fact that the provirus contains a pBR322-derived origin of replication and the Neo^r gene which confers resistance to kanamycin in bacteria. Chromosomal DNA was restricted with *Nhe*I and subsequently circularized. After transformation to *E. coli*, a clone which had the expected restriction pattern was isolated and designated pBB10c37.1. As the *Nhe*I site is located in the extreme 5′ part of the LTR, a plasmid is generated which contains virtually the complete 5′ LTR (Fig. 1A). After transfection into psi-2 cells, this LTR drives transcription and provides the polyadenylation signal as well. BB10 virus-containing medium was isolated 3 days posttransfection and used to infect B77 cells. A total of 23 G418^r clones were isolated and stained for β-gal. After staining, most colonies were mosaic with respect to β-gal expression. The spectrum of phenotypes in this group of clones was equivalent to that of the colonies derived after infection with the original stocks (Fig. 6). In addition, DNA sequence analysis of the LTR-containing *Nhe*I-*Sma*I fragment from plasmid pBB10c37.1 revealed a sequence identical to that of parental clone (data not shown). From these results, we conclude that the BAG provirus in BB10 is not inactivated as the result of mutation of vector sequences. Therefore, we favor the hypothesis that the shutdown of expression is governed by regulatory elements in the flanking host cell genome.

TABLE 1. Activity of Mo-MuLV LTR promoter linked to the CAT gene in B77 and its derivative cell lines and in the EC cell line P19

Cell line	% Acetylation			Ratio (MuLV/RSV)
	pBL-CAT5	pRSV-CAT	pMuLV-CAT	
B77	0.06	13.9	34.1	2.45
BB10	0.04	12.1	26.2	2.17
BB16	0.06	4.2	10.0	2.38
P19	0.06	1.9	0.2	0.11

DISCUSSION

We studied the expression from the Mo-MuLV LTR promoter in murine fibroblast cell line B77 infected with the

Downloaded from jvi.asm.org by on September 14, 2019

retroviral vector BAG (37). We showed that cell lines containing a single integrated copy of the vector provirus are heterogeneous in their expression of the reporter, *E. coli*-derived β-gal. In 60% of the G418^r colonies, more than 15% of the cells were β-gal negative (Fig. 1). This frequency is too high to be explained by mutational events but rather suggests an epigenetic mechanism. Treatment of β-gal-negative cells with the DNA-demethylating agent azaC resulted in resurrection of β-gal activity (Fig. 3A). However this azaC-induced reactivation was temporary (Fig. 3B); removal of azaC from the culture medium resulted in a gradual decline of the fraction of β-gal-expressing cells. From these data, we inferred that an epigenetic mechanism, viz., the de novo methylation of C residues, is responsible for shutdown of expression. This hypothesis was supported by the fact that the *Sma*I site in part of the DNA molecules isolated from BB10 cells, but not from BB16 cells, was refractory to digestion with *Sma*I, which is sensitive to methylation of its recognition site (38). The involvement of methylation of C residues in transcription repression is well established. Methylated LTRs are inefficient as promoters in murine fibroblast cell lines, whereas unmethylated LTRs are fully functional (19, 31, 46, 50; reviewed in reference 55).

Several other groups have studied the stability of Mo-MuLV-driven gene expression in fibroblast cell lines. It has been shown that such a loss of proviral gene expression can result from chromosome instability and subsequent loss of the vector provirus (49). However, in our case chromosome loss is not involved in the shutdown of expression as the provirus is obviously retained. In an extensive study, Xu and co-workers (60) assessed the stability of retroviral vectors after integration and observed, as in our study, shutdown of provirus expression without apparent alterations of vector sequences. In contrast to our results, these researchers were unable to restore expression by azaC treatment in their cells. However, in their study, the cells were grown in azaC-containing medium for only 24 h. In our hands, a 24-h treatment with azaC resulted only in a moderate reactivation of expression, whereas at least 3 days are required for a maximum effect (data not shown).

The mechanism by which the shutdown is accomplished is still obscure. In our study, we showed that demethylation alone is sufficient to reactivate transcription of the provirus in murine fibroblast cell lines. This suggests that the presence of methylated C residues in the 5′ LTR of the provirus constitutes the only block to transcription. Transfection experiments in which the MuLV LTR was linked to the CAT gene (Table 1) demonstrated that the MuLV LTR enhancer-promoter is potentially fully functional in these cells. In contrast, provirus expression in EC cells can only be restored by azaC treatment after differentiation of the cells (36), and extensive methylation of inactive proviruses can only be observed after a lag period of many days (12; reviewed in reference 55). This suggests that the extensive methylation in these EC cells is a consequence of the inactivation rather than an initial step. Our data suggest that the inactivation by way of methylation of the Mo-MuLV LTR is not necessarily the same as the transcription repression mechanism that is operating in undifferentiated EC cells. The proteins that are involved in maintaining the methylation patterns have been identified, but little is known about the factors responsible for de novo methylation (reviewed in reference 1).

The frequency of the epigenetic shutdown of expression varies between different cell clones. We demonstrated that the inactivation of the Mo-MuLV LTR in murine fibroblast cell lines is not due to a high incidence of mutations that might increase the susceptibility to methylation. Therefore, we conclude that the frequency of inactivation is modulated by positional effects. Retrovirus integration into the host cell genome is not a random process. Apart from sequence-independent and apparently randomly distributed integration sites, a limited number of strongly preferred sites is used (44). Such sites are frequently located in transcriptionally active regions of the genome (33, 42) or near DNase I-hypersensitive regions (39, 56). The vicinity of cellular regulatory elements might influence the function of integrated viral counterparts. Differences in the onset of viremia in the various transgenic *Mov* mice carrying complete copies of the Mo-MuLV provirus in their germ line have been attributed to *cis*-acting effects of flanking chromosomal DNA (6, 22, 23). However, it cannot be excluded that small mutations in the proviruses contribute to this variability. In undifferentiated EC cells, the block of transcription can be overcome by the proximity of cellular enhancers (2, 51). Likewise, integration into silent regions of the genome may result in down-regulation of the provirus.

We showed that inactivation of the provirus can occur in cells in which the LTR enhancer-promoter is potentially fully functional. This implies that the shutdown mechanism is dominant over the viral regulatory elements, i.e., that the regulatory elements in the provirus LTR are not sufficient for sustained expression of the provirus in a position-independent manner. Recently, regulatory elements (termed dominant control regions) have been identified that can confer cell-type specificity and position independence to linked genes after transfer into the mouse germ line (16, 17). We conclude that the Mo-MuLV LTR does not contain sequences that can act as a dominant control region in murine fibroblast cell lines. As a consequence, the Mo-MuLV LTR enhancer-promoter might not be sufficient to ensure sustained high-level expression in the absence of selective pressure, even in cells in which the enhancer-promoter is potentially fully functional. It could be the position-dependent inactivation that is responsible for the clonal variation in expression often observed, e.g., in the progeny of individual stem cells after retrovirus-mediated gene transfer (3, 4, 24, 32, 54, 59).

Downloaded from jvi.asm.org by on September 14, 2009

ACKNOWLEDGMENTS

We thank Bob Scholte (Erasmus University, Rotterdam, The Netherlands) and Robin Wielders for help in some of the experiments, and Dinko Valerio (Institute for Applied Radiobiology and Immunology, Rijswijk, The Netherlands), Christiaan Karreman (Department of Pediatrics, Academic Hospital, Leiden, The Netherlands), Ingeborg Meyer, and A. G. Jochemsen for stimulating and helpful discussions. We also thank Conny Cepko (Harvard Medical School, Boston, Mass.) for providing pBAG plasmid DNA.

This work was supported in part by the Netherlands Organization for the Advancement of Pure Research (N.W.O.) through the Foundation of Chemical Research in the Netherlands (S.O.N.) and by the Stichting Hemophilia (Dutch Hemophilia Foundation).

REFERENCES

1. **Adams, R. L. P.** 1990. DNA methylation: the effect of minor bases on DNA-protein interactions. Biochem. J. 265:309–320.
2. **Barklis, E., R. C. Mulligan, and R. Jaenisch.** 1986. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. Cell 47:391–399.
3. **Belmont, J. W., J. Henkel-Tigges, S. M. W. Chang, K. Wager-Smith, R. E. Kellems, J. E. Dick, M. C. Magli, R. A. Phillips, A. Bernstein, and C. T. Caskey.** 1986. Expression of human adenosine deaminase in murine haematopoietic progenitor cells

following retroviral transfer. Nature (London) 322:385–387.
4. Belmont, J. W., G. R. MacGregor, K. Wager-Smith, F. A. Fletcher, K. A. Moore, D. Hawkins, D. Villalon, S. M.-W. Chang, and C. T. Caskey. 1988. Expression of human adenosine deaminase in murine hematopoietic cells. Mol. Cell. Biol. 8:5116–5125.
5. Brands, J. H. G. M., J. A. Maassen, F. J. Van Hemert, R. Amons, and W. Möller. 1986. The primary structure of the α-subunit of human elongation factor 1. Eur. J. Biochem. 155:167–171.
6. Chumakov, I., H. Stuhlmann, K. Harbers, and R. Jaenisch. 1982. Cloning of two genetically transmitted Moloney leukemia proviral genomes: correlation between biological activity of cloned DNA and viral genome activation in the animal. J. Virol. 42:1088–1098.
7. Coffin, J. 1984. Endogenous retroviruses, p. 1109–1203. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132:6–13.
9. Feuer, G., M. Taketo, R. C. Hanecak, and H. Fan. 1989. Two blocks in Moloney murine leukemia virus expression in undifferentiated F9 embryonal carcinoma cells as determined by transient expression assays. J. Virol. 63:2317–2324.
10. Franz, T., F. Hilberg, B. Sellinger, C. Stocking, and W. Ostertag. 1986. Retroviral mutants efficiently expressed in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 83:3292–3296.
11. Friedmann, T. 1989. Progress toward human gene therapy. Science 244:1275–1281.
12. Gautsch, J. W., and M. C. Wilson. 1983. Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature (London) 301:32–37.
13. Gorman, C. 1985. High efficiency gene transfer into mammalian cells, p. 143–190. *In* D. M. Glover (ed.), DNA cloning: a practical approach, vol. II. IRL Press, Oxford.
14. Gorman, C. M., P. W. J. Rigby, and D. P. Lane. 1985. Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell 42:519–526.
15. Graham, F. L., and A. J. van der Eb. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology 52:456–467.
16. Greaves, D. R., F. D. Wilson, G. Lang, and D. Kioussis. 1989. Human CD2 3′-flanking sequences confer high-level, T-cell-specific, position-independent gene expression in transgenic mice. Cell 56:979–986.
17. Grosveld, G., G. Blom van Assendelft, D. R. Greaves, and G. Kollias. 1987. Position-independent high-level expression of the human β-globin gene in transgenic mice. Cell 51:975–985.
18. Guild, B. C., M. H. Finer, D. E. Housman, and R. C. Mulligan. 1988. Development of retrovirus vectors useful for expressing genes in cultured murine embryonal cells and hematopoietic cells in vivo. J. Virol. 62:3795–3801.
19. Harbers, K., A. Schnieke, H. Stuhlmann, D. Jahner, and R. Jaenisch. 1981. DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning. Proc. Natl. Acad. Sci. USA 78:7609–7613.
20. Hilberg, F., C. Stocking, W. Ostertag, and M. Grez. 1987. Functional analysis of a viral host-range mutant: altered long terminal repeat sequences allow expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 84:5232–5236.
20a. Hoeben, R. C., and A. A. J. Migchielsen. Unpublished data.
21. Hoeben, R. C., R. C. M. van der Jagt, F. Schoute, N. van Tilburg, M. P. Verbeet, E. Briët, H. van Ormondt, and A. J. van der Eb. 1990. Expression of functional factor VIII in primary human skin fibroblasts after retrovirus-mediated gene transfer. J. Biol. Chem. 265:7318–7323.
22. Jaenisch, R., and D. Jahner. 1984. Methylation, expression and chromosomal position of genes in mammals. Biochim. Biophys. Acta 782:1–9.
23. Jaenisch, R., D. Jahner, P. Nobis, I. Simon, J. Lohler, K. Harbers, and M. Grotkopp. 1981. Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell 24:519–529.
24. Kaleko, M., J. V. Garcia, W. R. A. Osborne, and A. D. Miller. 1990. Expression of human adenosine deaminase in mice after transplantation of genetically-modified bone marrow. Blood 75:1733–1741.
25. Li, L. H., E. J. Olin, H. H. Buskirk, and L. M. Reineke. 1970. Cytotoxicity and mode of action of 5-azacytidine on L1210 leukemia. Cancer Res. 30:2760–2769.
26. Linney, E., B. Davis, J. Overhauser, E. Chao, and H. Fan. 1984. Non-function of a Moloney murine leukemia virus regulatory sequence in F9 embryonal carcinoma cells. Nature (London) 308:470–472.
27. Loh, T. P., L. L. Sievert, and R. W. Scott. 1987. Proviral sequences that restrict retroviral expression in mouse embryonal carcinoma cells. Mol. Cell. Biol. 7:3775–3784.
28. Loh, T. P., L. L. Sievert, and R. W. Scott. 1988. Negative regulation of retrovirus expression in embryonal carcinoma cells is mediated by an intragenic domain. J. Virol. 62:4086–4095.
29. McIvor, R. S., M. J. Johnson, A. D. Miller, S. Pitts, S. R. Williams, D. Valerio, D. W. Martin, Jr., and I. M. Verma. 1987. Human purine nucleoside phosphorylase and adenosine deaminase: gene transfer into cultured cells and murine hematopoietic stem cells by using recombinant amphotropic retroviruses. Mol. Cell. Biol. 7:838–846.
30. Miller, A. D., R. J. Eckner, D. J. Jolly, T. Friedman, and I. M. Verma. 1984. Expression of a retrovirus encoding human HPRT in mice. Science 225:630–632.
31. Montandon, P.-E., F. Montandon, and H. Fan. 1982. Methylation state and DNase I sensitivity of chromatin-containing Moloney murine leukemia virus DNA in exogenously infected mouse cells. J. Virol. 44:475–486.
32. Moore, K. A., F. A. Fletcher, D. K. Villalon, A. E. Utter, and J. W. Belmont. 1990. Human adenosine deaminase expression in mice. Blood 75:2085–2092.
33. Mooslehner, K., U. Karis, and K. Harbers. 1990. Retroviral integration sites in transgenic *Mov* mice frequently map in the vicinity of transcribed DNA regions. J. Virol. 64:3056–3058.
34. Mummery, C. L., A. Feijen, P. T. Van de Saag, C. E. Van den Brink, and S. W. De Laat. 1985. Clonal variants of differentiated P19 embryonal carcinoma cells exhibit epidermal growth factor receptor kinase activity. Dev. Biol. 109:402–410.
35. Niwa, O. 1985. Suppression of the hypomethylated Moloney leukemia virus genome in undifferentiated teratocarcinoma cells and inefficiency of transformation by a bacterial gene under control of the long terminal repeat. Mol. Cell. Biol. 5:2325–2331.
36. Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara. 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell 32:1105–1113.
37. Price, J., D. Turner, and C. Cepko. 1987. Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer. Proc. Natl. Acad. Sci. USA 84:156–160.
38. Roberts, R. 1990. Restriction enzymes and their isoschizomers. Nucleic Acids Res. 18:S2331–S2365.
39. Rohdewohld, H., W. Weiher, R. Reik, R. Jaenisch, and M. Breindl. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia virus integration sites map near DNase I-hypersensitive sites. J. Virol. 63:336–343.
40. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
41. Sanes, J. R., J. L. R. Rubenstein, and J.-F. Nicolas. 1986. Use of a retrovirus to study post-implantation cell lineage in mouse embryos. EMBO J. 5:3133–3142.
42. Scherdin, U., K. Rhodes, and M. Breindl. 1990. Transcriptionally active genome regions are preferred integration sites for retrovirus integration. J. Virol. 64:907–912.
43. Schrier, P. I., R. Bernards, R. T. M. J. Vaessen, A. Houweling, and A. J. van der Eb. 1983. Expression of class 1 major histocompatibility antigens switched off by highly oncogenic

Downloaded from pubs.asm.org by on September 14, 2009

adenovirus 12 in transformed rat cells. Nature (London) **305:** 771–775.
44. **Shih, C. C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53:**531–537.
45. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliff.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:**543–548.
46. **Simon, D., H. Stuhlmann, D. Jahner, H. Wagner, E. Werner, and R. Jaenisch.** 1983. Retrovirus genomes methylated by mammalian but not by bacterial methylase are non-infectious. Nature (London) **304:**275–276.
47. **Soriano, P., R. D. Cone, R. C. Mulligan, and R. Jaenisch.** 1986. Tissue-specific and ectopic expression of genes introduced in transgenic mice by retroviruses. Science **234:**1409–1413.
48. **Stein, B., H. J. Ramsdorf, A. Steffen, M. Litfin, and P. Herrlich.** 1989. UV-induced DNA damage is an intermediate in the UV-induced expression of human immunodeficiency virus type 1, collagenase, *c-fos*, and metallothionein. Mol. Cell. Biol. **9:**5169–5181.
49. **Stone, J. C., N. A. Dower, and L. Siminovitch.** 1986. Stability of retrovirally transduced markers in a rat cell line. Somat. Cell Mol. Genet. **12:**575–583.
50. **Stuhlmann, H., D. Jahner, and R. Jaenisch.** 1981. Infectivity and methylation of retroviral genomes is correlated with expression in the animal. Cell **26:**221–232.
51. **Taketo, M., and M. Tanaka.** 1987. A cellular enhancer of retrovirus gene expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA **84:**3748–3752.
52. **Tsukiyama, T., O. Niwa, and K. Yokoro.** 1989. Mechanism of suppression of the long terminal repeat of Moloney leukemia virus in mouse embryonal carcinoma cells. Mol. Cell. Biol. **9:**4670–4676.
53. **Valerio, D., M. F. W. Einerhand, P. M. Wamsley, T. A. Bakx, C. L. Li, and I. M. Verma.** 1989. Retrovirus-mediated gene transfer into embryonal carcinoma and hemopoietic stem cells: expression from a hybrid long terminal repeat. Gene **84:**419–427.
54. **Valerio, D., T. P. Visser, G. Wagemaker, A. J. van der Eb, and D. W. Van Bekkum.** 1986. Introduction of human ADA sequences into mouse hematopoietic cells, p. 335–355. *In* J. Vossen and C. Griscilli (ed.), Progress in immunodeficiency research and therapy, vol. 2. Elsevier Biomedical Press, Amsterdam.
55. **Varmus, H., and R. Swanstrom.** 1985. Replication of retroviruses, p. 75–134. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Supplement and appendices. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
56. **Vijaya, S., D. L. Steffen, and H. L. Robinson.** 1986. Acceptor sites for retroviral integration map near DNase I-hypersensitive sites in chromatin. J. Virol. **60:**683–692.
57. **Weiher, H., E. Barklis, W. Ostertag, and R. Jaenisch.** 1987. Two distinct sequence elements mediate retroviral gene expression in embryonal carcinoma cells. J. Virol. **61:**2742–2746.
58. **Williams, D. A., S. H. Orkin, and R. C. Mulligan.** 1986. Retrovirus-mediated transfer of human adenosine deaminase gene sequences into cells in culture and into murine hematopoietic cells in vivo. Proc. Natl. Acad. Sci. USA **83:**2566–2570.
59. **Wilson, J. M., O. Danos, M. Grossman, D. H. Raulet, and R. C. Mulligan.** 1990. Expression of human adenosine deaminase in mice reconstituted with retrovirus-transduced haematopoietic cells. Proc. Natl. Acad. Sci. USA **87:**439–443.
60. **Xu, L., J.-K. Yee, J. A. Wolff, and T. Friedmann.** 1989. Factors affecting long-term stability of Moloney murine leukemia virus-based vectors. Virology **171:**331–341.
61. **Yoshida, S., M. Yamada, and S. Masaki.** 1970. Inhibition of DNA polymerase-α and -β of calf thymus by 1-β-D-arabinofuranosyl-cytosine-5'-triphosphate. Biochim. Biophys. Acta **447:** 144–150.

Downloaded from jvi.asm.org by on September 14, 2009

TAB 10

MOLECULAR AND CELLULAR BIOLOGY, Apr. 1986, p. 1141–1147
0270-7306/86/041141-07$02.00/0
Copyright © 1986, American Society for Microbiology
Vol. 6, No. 4

Variable Stability of a Selectable Provirus after Retroviral Vector Gene Transfer into Human Cells

DOUGLAS J. JOLLY,[1] RANDALL C. WILLIS,[2] AND THEODORE FRIEDMANN[1]*

Departments of Pediatrics[1] and Medicine,[2] University of California at San Diego School of Medicine, La Jolla, California 92093

Received 12 August 1985/Accepted 30 December 1985

Human lymphoblasts deficient in the enzyme hypoxanthine-guanine phosphoribosyltransferase (HPRT) were infected with an amphotropic helper-free retroviral vector expressing human HPRT cDNA. The stability and expression of the HPRT provirus in five cell lines with different proviral integration sites were examined by determining HPRT mutation and reversion frequencies and by blot hybridization studies. Mutation to the HRPT-negative phenotype occurred at frequencies of approximately 4×10^{-5} to 3×10^{-6} per generation. Most mutations in each of the five cell lines were associated with partial or complete deletions or rearrangements of the provirus. Several mutants retained a grossly intact HPRT provirus, and in one such mutant HPRT shutdown resulted from a revertable epigenetic mechanism that was not associated with global changes in proviral methylation. Therefore, mutation and shutdown of the HPRT provirus in human lymphoblasts result from mechanisms similar to those reported for several other avian and mammalian replication-competent retroviruses.

Defective forms of eucaryotic retroviruses have recently become useful for studying the transfer of genetic information into mammalian cells because of their very high efficiency of infection, their large capacity for added sequences, and their fairly well-understood mechanisms of infection, integration, and gene expression in recipient cell genomes (15, 16, 21, 24, 28, 32, 34, 35, 43, 44). However, it is known that retroviral infection of eucaryotic cells can be accompanied by proviral instability, as shown by the occasionally high rates of reversion and shutdown of gene expression and by changes in proviral structure, including deletions, point mutations, losses of entire proviruses, and changes in proviral methylation (1, 3, 6, 7, 9–12, 17, 19, 22, 23, 25, 26, 30, 31, 39–42, 47). It has not been clear how many of these instabilities are due to error-prone replication and how many are due to nonreplicative mechanisms. It is known that rat cells expressing an avian sarcoma virus revert to the nontransformed phenotype by several mechanisms, including loss of the entire provirus and nonconditional mutations in the *src* transforming gene (41). However, similar direct studies of the stability of nontransforming proviruses have not been extensive due to the paucity of easily selectable proviral gene products.

We investigated the stability of proviral organization and expression by using nonproducer, hypoxanthine-guanine phosphoribosyltransferase (HPRT)-deficient human lymphoblast cell lines genetically transformed to the $HPRT^+$ phenotype by infection with an HPRT retroviral vector (21). In this study we took advantage of the facility and efficiency of selection both for and against cells expressing HPRT activity in tissue culture (13, 46), thereby allowing the detection and characterization of rare genetic events that affect the expression of the HPRT gene. Evidence from Southern blots, mutation and reversion frequencies, and enzyme activities indicated that in human HPRT-deficient lymphoblasts infected with an HPRT retroviral vector, integrated HPRT proviruses are relatively stable, showing frequencies of mutation from $HPRT^+$ to $HPRT^-$ ranging from 3×10^{-6} to 4×10^{-5} per generation. The mechanisms of mutation to the $HPRT^-$ phenotype vary in cell lines carrying proviruses integrated at different sites and include a variety of deletions, rearrangements, and epigenetic events. These findings may be of use in developing an understanding of the fate and expression of defective retroviral vectors in human and other mammalian cells.

MATERIALS AND METHODS

Cells and culture. Human Lesch-Nyhan HPRT-deficient lymphoblasts (46) transformed by Epstein-Barr virus were grown in RPMI 1640 medium containing 10% fetal calf serum at 37°C in 5% CO_2. The enzyme-deficient cells were infected with a transmissible HPRT vector and were selected and characterized as described previously (46). Independently isolated $HPRT^+$ cell lines 1, 9, 11, 13, and 16 were chosen for further characterization and were designated the parent cell lines in this study.

Selection of $HPRT^-$ cells was performed with 20 μM 6-thioguanine, and the selection medium for $HPRT^+$ cells contained 10^{-4} M hypoxanthine, 2.5×10^{-7} M aminopterin, and 5×10^{-4} M thymidine (HAT) added to the medium described above. Mutation frequencies were measured by growing cells in HAT medium, withdrawing them from HAT medium for periods of 1 to 100 days, and plating them in 0.3% soft agar containing 2 μM thioguanine. The number of resistant colonies, corrected for plating efficiency, was determined after 10 to 14 days, and the mutation frequency per generation was calculated from the initial slope of the curve correlating the time without selection with the number of mutants. Reversion from $HPRT^-$ to $HPRT^+$ was measured by analogous procedures, using HAT selective media.

Parental cell lines were determined to be nonproducers (i.e., not infected with a competent retrovirus) by growing the infected $HPRT^+$ cells to a density of 0.5×10^6 cells per ml, replacing the medium with fresh medium containing 4 μg of Polybrene per ml, harvesting the medium 16 h later, and testing for the presence of HPRT infectious units on a

* Corresponding author.

Downloaded from mcb.asm.org by on September 14, 2009

TABLE 1. HPRT levels and mutation and reversion frequencies of HPRT-infected cell lines

Cell line	HPRT level (% of wild type)	Mutation to thioguanine resistance		HAT^r reversion		Second-cycle mutation to thioguanine resistance	
		Frequency	HPRT level (% of wild type)[a]	Frequency	HPRT level (% of wild type)	Frequency	HPRT level (% of wild type)[a]
1	20	3×10^{-6}	<0.5[b]	$<10^{-7}$			
9	11	3×10^{-5}	<0.5[c]	$<10^{-7}$			
11	18	3×10^{-6}	<0.5[d]	$<10^{-7}$			
13	23	4×10^{-5}	<0.5[d]	$<10^{-7}$			
16	5	1.5×10^{-5}	<0.5[d]	1.3×10^{-4}	4[d]	3.3×10^{-3}[e]	<0.5

[a] The levels were indistinguishable from the level in the original Lesch-Nyhan parent cell line.
[b] Determined by using two clones; HPRT activities were measured individually.
[c] Determined by using six clones; HPRT activities were measured individually.
[d] Determined by using five clones; HPRT activities were measured individually.
[e] Determined by using cell line 16b revertant 3.

recipient lawn of HPRT-deficient human SV40-transformed Lesch-Nyhan (LNSV) cells or rat 208F cells. Two cell lines (cell lines 16 and 11) were infected with competent 1504A virus (21) and grown for 2 weeks before the supernatants were assayed for HPRT virus production.

For mutation and reversion studies, colonies were plated after no more than 1 day out of HAT or thioguanine selective conditions in order to minimize the possibility of picking sibling clones.

HPRT activities. Cells were grown to mid- to late-log phase and harvested, and cell extracts were prepared and assayed for HPRT activity as described previously (14, 46), using [^{14}C]-hypoxanthine as a substrate. HPRT activities were normalized to levels of adenosine phosphoribosyl transferase (46).

DNA and RNA preparation, restriction enzyme digests, and Southern blots. DNA and RNA were prepared as described previously (21). Restriction enzymes *Sst*I, *Sac*I, *Pvu*II, *Eco*RI, *Bam*HI, *Hpa*II, and *Msp*I were purchased from Amersham Corp., New England BioLabs, Inc., or Bethesda Research Laboratories and were used as recommended by the manufacturers. Agarose gel analysis of enzyme-digested DNA and transfer of DNA to nitrocellulose filters by Southern blotting (33) were performed as described previously (14, 46). Hybridization with a nick-translated probe (18) was performed in 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7) at 68°C. The filters were washed initially with 2× SSC and finally with 0.2× SSC at 68°C. Hybridization probes were purified free of plasmid vector sequences by agarose gel electrophoresis after restriction enzyme digestion and nick translated to a specific activity of approximately 2×10^8 cpm/μg. The following DNA probes were used: for HPRT, the *Pst*I-to-*Rsa*I fragment from cDNA plasmid p4aA8 containing the full coding portion of the cDNA and a portion of the untranslated 3′ region (bases 1 to 890) (14); for the long terminal repeat (LTR), the double LTR fragment cloned with R1 linkers in pLTR10 (46).

RESULTS

Five previously described (46) human, Epstein-Barr virus-transformed, nonproducer lymphoblast cell lines derived from a Lesch-Nyhan patient and made HAT resistant by infection with a transmissible HPRT retroviral vector were tested for their frequency of mutation to thioguanine resistance (i.e., HPRT$^-$). These cell lines had HPRT activity ranging from 5 to 23% of the wild type levels, and all contained one detectable integrated copy of the HPRT provirus (46). Their properties, mutation frequencies, and HPRT enzyme levels are shown in Table 1. Thioguanine was used as a selective agent because it is known to give few, if any, HPRT$^+$ thioguanine-resistant cells, in contrast to azaguanine (27, 29). Table 1 also shows the frequencies with which the thioguanine-resistant (HPRT$^-$) mutants reverted to the HAT-resistant HPRT$^+$ phenotype.

The mutation rate of control cell line WI-L2 during selection of the HPRT-negative phenotype with thioguanine has been reported to be anywhere from less that 2×10^{-7} to 10^{-6} per cell generation (4, 5, 36–38; Willis, unpublished data), values that depend on the history of the cell line, the cell density, the concentration of the selective agent, and other culture parameters. Using the conditions described here, workers in our laboratories have found that the mutation frequency of cell line WI-L2 is approximately 10^{-6} per cell generation. The mutation frequencies for the five parent HPRT-positive Lesch-Nyhan lymphoblast transformants ranged from 4×10^{-5} to 3×10^{-6} per generation, values that are approximately equal to, but probably somewhat higher than, the mutation frequency reported for the HPRT locus in WI-L2 cells and for a variety of other single-copy genes in tissue culture cells.

In order to determine the mechanism of the mutation in these cells, we examined the structure of the HPRT provirus by Southern blotting DNAs from the thioguanine-resistant cells. The organization of the HPRT provirus in HPRT-deficient cells infected by the transmissible HPRT vector is shown in Fig. 1. Digestion of DNAs from the HPRT-infected parent cells with *Sac*I, its isoschizomer (*Sst*I), or other enzymes that cut an intact, unrearranged HPRT provirus only in the LTRs would be expected to reveal a single 4-kilobase (kb) band with an HPRT probe, while digestion with *Bam*HI should reveal one band of variable size (but always larger than 2.6 kb by an amount dependent on the site of proviral integration and the location of the nearest *Bam*HI site in the flanking host sequences). Similarly, digestion of cellular DNA with *Eco*RI and hybridization with an HPRT probe should reveal one band of variable size, always larger than 4.6 kb, since there are no *Eco*RI sites within the provirus. All of these patterns, of course, are superimposed on fragments derived from the endogenous, nonfunctional HPRT gene sequences known to be present in these cells (46).

With Moloney murine sarcoma virus LTR probes that did not hybridize to a significant extent with human DNA under our standard conditions of stringency, *Sst*I-digested DNA should have revealed the 4-kb band identical to the band visualized with the HPRT probe, in addition to another major band of variable size. Hybridization with the same LTR probes to blots of *Bam*HI-digested DNA should have

Downloaded from mcb.asm.org by on September 14, 2009

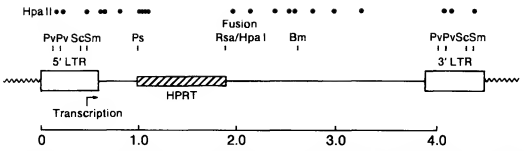


FIG. 1. Map of the HPRT provirus in infected cells. The open boxes represent the LTR sequences, and the cross-hatched box represents the HPRT sequence. The *Hpa*II (and hence *Msp*I) restriction enzyme sites are indicated. The other sites include *Pvu*II (Pv), *Sac*I or *Sst*I (Sc), *Sma*I (Sm), *Pst*I (Ps), and *Bam*HI (Bm) sites. The scale is in kilobase pairs. The cellular DNA sequences immediately adjacent to the 5′ and 3′ LTR sequences are indicated by wavy lines.

revealed two fragments of variable size, one of which corresponded to the same fragment detected with an HPRT probe and one of which was unique. Blots of cellular DNA digested with *Eco*RI hybridized to the LTR probe should have revealed one band identical to the band revealed by the HPRT probe and larger than 4.6 kb.

Combinations of restriction enzymes and hybridization probes were used to characterize the structure of the HPRT provirus in the five parent cell lines grown with and without continued HAT selection in mutants and revertants.

Cell line 13. In six HPRT-negative mutants derived from tetraploid (46) parent cell line 13 (three mutants derived from each of two separate experiments), all of the HPRT and LTR proviral sequences were lost, and in the one mutant line which we examined by karyotype analysis the proviral loss was accompanied by the loss of chromosome 4 (data not shown). No proviral rearrangements were detected in parent cells after prolonged growth for 22 generations without selection.

Cell line 9. The results obtained with tetraploid cell line 9 are shown in Fig. 2. Mutants 9a, 9b, and 9c and mutants 9d, 9e, and 9f were derived from two separate experiments. As was the case with cell line 13, there was no detectable difference between the DNAs from cells cultured with HAT selection and the DNAs from cells cultured without HAT selection for 22 generations (Fig. 2, lanes 2 and 3). All six HPRT$^-$ mutants (mutants 9a, 9b, 9c, 9d, 9e, and 9f) (Fig. 2, lanes 4 through 9) had lost most or all of a 7.5-kb *Bam*HI fragment containing the HPRT sequences found in the parent cell lines (lanes 2 and 3), leaving only the endogenous HPRT gene pattern (lane 1).

The same filter containing *Bam*HI-digested DNA hybridization to the LTR probe showed that two of the six revertants had also lost the LTR sequences (Fig. 2, lanes 13 and 16). Three of the mutants (mutants 9a, 9d, and 9f) (lanes 12, 15, and 17) produced LTR patterns which were different from those of the parent cells but very similar to each other, retaining the parental HPRT-negative, LTR-positive, 13-kb *Bam*HI fragment and also showing two new *Bam*HI fragments (approximately 15 and 5 kb). Mutant 9c (lane 6) retained only the 15-kb LTR-positive fragment (lane 14). Since the same filter was used for hybridization to both the

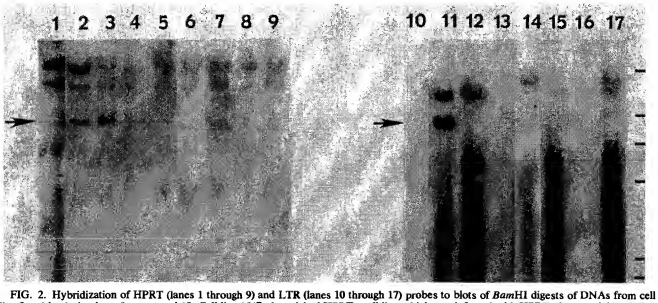


FIG. 2. Hybridization of HPRT (lanes 1 through 9) and LTR (lanes 10 through 17) probes to blots of *Bam*HI digests of DNAs from cell line 9 and its derivatives. Lanes 1 and 10, Cell line 1547, the original HPRT$^-$ cell line which was infected with HPRT virus to yield cell line 9; lane 2, cell line 9 maintained for 22 weeks without HAT selection; lanes 3 and 11, cell line 9 maintained with HAT selection; lanes 4 through 9 and 12 through 17, HPRT$^-$ mutants 9a, 9b, 9c, 9d, 9e, and 9f, respectively. The arrows indicate the approximately 7.5-kb band in cell line 9 (lanes 2, 3, and 11) which was apparent in the HPRT and LTR hybridizations but was not detectable in the HPRT$^-$ mutants (lanes 4 through 9 and 12 through 17). The markers indicated on the right are λ *Hin*dIII fragments (23.6, 9.6, 6.6, 4.3, 2.3, and 2.0 kb from top to bottom).

Downloaded from mcb.asm.org by on September 14, 2009

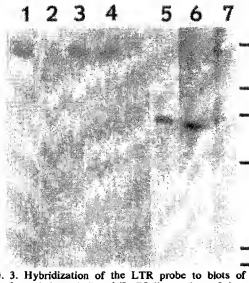


FIG. 3. Hybridization of the LTR probe to blots of *Bam*HI digests (lanes 1 through 4) and *Eco*RI digests (lanes 5 through 7) of DNAs from cell line 1 and its derivatives. Lanes 1 and 5, Parent cell line 1 maintained with HAT selection; lanes 2 and 6, HPRT$^-$ mutant cell line 1e; lanes 3 and 7, HPRT$^-$ mutant cell line 1f; lane 4, parent cell line 1 maintained without HAT selection for 22 weeks. The markers indicated on the right are λ *Hin*dIII fragments (see the legend to Fig. 2).

LTR and the HPRT probes, the low-molecular-weight smear found in several mutants, as well as in the parent cell lines, cannot be explained by sample degradation or variations in sample loading.

Cell line 1. The results obtained with diploid cell line 1 are shown in Fig. 3. There was no detectable difference in proviral structure between the parent cell line maintained in HAT medium and the cell line grown for 22 generations without selection for the HPRT$^+$ phenotype (Fig. 3, lanes 1 and 4). We characterized two HPRT-negative clones, clones 1e and 1f, by Southern blotting with both the LTR (Fig. 3) and HPRT probes (data not shown). Mutant 1f (lanes 3 and 7) was indistinguishable from its parent, while the major band in mutant 1e in both *Eco*RI and *Bam*HI digests (lanes 2 and 6) migrated reproducibly slightly faster than the major band of the other revertant or of the parent line grown with or without HAT selection. Hybridization of the HPRT probe to *Eco*RI digests showed that HPRT sequences were retained in both mutants and migrated in the same position as the LTR-hybridizing band.

Cell line 11. The results obtained with diploid cell line 11 are shown in Fig. 4. Mutants 11a, 11b, and 11c (Fig. 4, lanes 2 through 4) and mutants 11e and 11f (lanes 5 and 6) were derived from two separate experiments. Southern blots of cellular DNAs from all five mutants digested with *Bam*HI (Fig. 4A) or with *Sst*I (Fig. 4B) and hybridized to the LTR probe showed patterns that were identical or nearly identical to each other but different from the pattern for the parent line (lane 1). All of the mutants were completely devoid of proviral HPRT sequences (data not shown). In *Sst*I digests, a fragment (Fig. 4B, arrow) which hybridized with an LTR probe but not with an HPRT probe was found in the position of one of the LTR bands in the parental cell line. We obtained similar results with Southern blots of cellular DNA digested with *Pvu*II, which also cut the HPRT proviral DNA only in the LTRs (Fig. 1) (data not shown). Such patterns are most easily explained by homologous recombination between the 5′ and 3′ LTRs, resulting in the excision of the single proviral *Sst*I or *Pvu*II fragment and the persistence of unaltered LTR fragments.

Cell line 16. Figure 5 shows the mutation and reversion frequencies for diploid line 16. Mutants 16a, 16b, and 16c (Fig. 5, lanes 3 through 5) and mutants 16e and 16f (lanes 6 and 7) were derived from two separate experiments. The same *Bam*HI digests were hybridized to the LTR (Fig. 5a) and HPRT (Fig. 5b) probes. The HPRT and LTR blots showed that the parent cell line (lanes 1 and 2) had an LTR-containing proviral band at 4.3 kb (Fig. 5a, arrow B) that also hybridized to the HPRT probe (Fig. 5b, arrow C). As expected (Fig. 1), the parent cell line also had a second LTR-hybridizing band, at approximately 11 kb. This arrangement was not detectably altered in two HPRT$^-$ mutants (lanes 4 and 5). However, in mutant 16a (lane 3), the upper LTR band was apparently increased in size, and in cell lines 16e and 16f (lanes 6 and 7) the HPRT-containing 4.3-kb band was completely lost and the upper band was reduced in size in at least one of these cell lines (lane 7).

HPRT$^-$ mutant 16b (lane 4) reverted to HAT resistance (Table 1), and the results of hybridization to digests of DNAs from several such HPRT$^+$ revertant lines (cell lines 16br1, 16br2, and 16br3) are shown in Fig. 5, lanes 8 through 10. In addition, the results obtained with DNA samples from four mutant cell lines derived from a second cycle of thioguanine selection of cell line 16br3 (lane 10) are shown in lanes 11 through 14. The HPRT and LTR blots showed no detectable rearrangement of the HPRT proviral sequences in these experiments after several sequential cycles between the HPRT-positive and -negative phenotypes.

Because of the stable Southern blot pattern and the very high frequency of HPRT phenotype switching, it seems likely that the modulation of HPRT gene expression in mutant 16b, its revertants, and second-cycle mutants is regulated by epigenetic events. Therefore, we examined the methylation pattern of the HPRT provirus after digestion with *Bam*HI and either *Hpa*II or *Msp*I (2) (see Fig. 1 for the locations of restriction enzyme sites in the proviral sequence). We found no significant differences in any of these

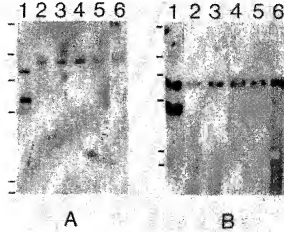


FIG. 4. Hybridization of the LTR probe to blots of digests of DNAs from cell line 11 and its derivatives. Lanes 1, Parent cell line 11; lanes 2 through 6, HPRT-negative mutant lines 11a, 11b, 11c, 11e, and 11f, respectively. (A) *Bam*HI digests. (B) *Sst*I digests. The arrow indicates the band that was conserved in the parent and mutant lines in the *Sst*I digests. The markers used were λ *Hin*dIII fragments (see the legend to Fig. 2).

Downloaded from mcb.asm.org by on September 14, 2009

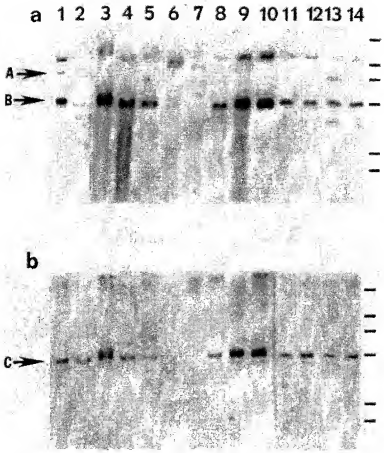


FIG. 5. Hybridization of LTR (a) and HPRT (b) probes to *Bam*HI digests of DNAs from cell line 16 and its derivatives. Lane 1, Cell line 16 preparation 2, which was made after extended growth with HAT selection (approximately 100 generations after infection of the parent HPRT$^-$ cell line with the HPRT virus); lane 2, initial DNA preparation of cell line 16 (approximately 25 generations after vector infection); lanes 3 through 7, HPRT$^-$ mutants 16a, 16b, 16c, 16e, and 16f, respectively; lanes 8 through 10 HPRT$^+$ revertants of mutant 16b (revertants 16br1, 16br2, and 16br3, respectively); lanes 11 through 14, HPRT$^-$ thioguanine mutants of revertant 16br3 (mutants 16br3rr1, 16br3rr2, 16br3rr3, and 16br3rr4, respectively). Arrow A indicates the extra LTR-hybridizing band at around 8 kb observed in parent cell line 16 after extended growth with HAT selection. The other arrows indicate the indistinguishable proviral bands of about 4 kb, which hybridized to the LTR (arrow B) and HPRT (arrow C) probes and were maintained in most of the cell lines. The markers used were λ *Hin*dIII fragments (see the legend to Fig. 2).

cell lines (data not shown), suggesting that the switch in these cells from HPRT gene expression to shutdown is not obviously correlated with global changes in the extent of methylation of the HPRT proviral sequences.

DISCUSSION

The HPRT gene in the form of transmissible retroviral vectors offers some major advantages as a model gene for the study of the stability of proviruses in mammalian cells. It provides an easily selectable, nontransforming provirus-derived gene function whose stability through repeated cycles of mutation and reversion of nonproducer cells permits distinction between replication-derived instability and proviral instability. Information obtained from such a selectable system ought to be relevant to other genes introduced into mammalian cells by similarly designed retroviral vectors, as well as to the parent retroviruses themselves.

The HPRT provirus which we used is relatively stable, showing frequencies of mutation to the HPRT$^-$ phenotype of approximately 4×10^{-5} to 3×10^{-6} per cell generation, a value slightly to moderately higher than the value obtained from the HPRT locus in the control WI-L2 cells. Nevertheless, the HPRT provirus is susceptible to several kinds of rearrangements or epigenetic events that accompany shutdown of proviral HPRT expression. In one of the parent cell lines (cell line 13) which was tetraploid before selection of HPRT-negative mutants, proviral loss was associated with loss of chromosome 4. We did not distinguish between the possibility that the provirus was eliminated through chromosome loss and the possibility that it was eliminated through coincidental but unrelated proviral excision. Similarly, whether chromosome loss is responsible for proviral loss in some of the mutants of the other tetraploid parent cell line (cell line 9) is not clear, but the persistence and modification of LTR sequences, together with the complete loss of HPRT sequences in several of the other mutants derived from cell line 9, indicate that proviral deletions and rearrangements occur in at least some mutants in this cell line. Less extensive changes seem to have occurred in the two mutants of parent cell line 1; one apparently resulted in a small detectable deletion, and the other resulted in a grossly intact provirus.

Downloaded from mcb.asm.org by on September 14, 2009

All of the mutants derived from cell line 11 have altered structures that are compatible with homologous recombination events between the proviral LTRs. Such a mechanism has been suggested for the reversion of the dilute coat color mutation in DBA/2J mice which results from the excision of an integrated mutagenic ecotropic leukemia virus (7) and for the reversion from nontransformed phenotype to transformed phenotype of a cell line in which an exogenous provirus interrupted a resident transforming gene (40). Similarly, some of the mutants of parent line 16 retained a grossly unaltered provirus, while others, like some of the mutants of cell line 9, lost the HPRT sequence but retained LTR sequences, albeit at times in rearranged form. At least one of the mutants (mutant 16b) obviously retained all of the sequences required for proviral HPRT expression since we were able to select $HPRT^+$ revertants (16br mutants) and second-cycle $HPRT^-$ mutants. The results of *Hpa*II and *Msp*I restriction enzyme digestion do suggest that DNA methylation is not a major factor responsible for the shutdown of viral gene expression in these cells (Jolly, unpublished data).

Our data indicate that a murine retrovirus-based amphotropic HPRT provirus integrated into human lymphoblasts after infection with a transmissible helper-free vector is relatively stable, but that rare cells are susceptible to proviral shutdown by a number of the same mechanisms which are known to modify avian and murine proviral structure and expression in other cells. Since only few of the isolated mutants contained grossly intact proviruses and since it was not possible (except for cell line 16b) to isolate any HAT-resistant reversions, we concluded that most mutations in this system involve deletions and rearrangements rather than single-base mutations. Since the parent cell lines in this study presumably differed from each other only in the site of proviral integration, we inferred that the various modes of proviral instability are influenced by the site of integration of the provirus or by flanking cellular sequences. Such a positional effect is not surprising in view of the well-known effect of integration site on some proviral expression (8, 11, 12).

Studies of the stability of integrated proviruses not only may be helpful in characterizing the general mechanisms of recombination involved in viral integration, but also are important components of studies to test the feasibility of retrovirus-mediated gene therapy models (20, 45). A reasonable requirement for such an approach to therapy is that the structure and expression of the genetic modification be stable. Our results indicate that the HPRT provirus is indeed reasonably stable in diploid human lymphoblasts, but that the several genetic and epigenetic mechanisms known to modify other proviruses also cause HPRT shutdown at detectable rates.

ACKNOWLEDGMENTS

This study was supported by Public Health Service grants GM28223 (to T.F.) and GM17702 (to. R.C.W.) from the National Institute of General Medical Sciences, by Public Health Service grant CA02290 from the National Cancer Institute, by the Weingart Foundation (T.F.), by the Keck Foundation (T.F.), by The Gould Family Foundation (T.F.), by The Kroc Foundation (T.F.), and by grant IN 93M from the American Cancer Society (to. D.J.J.).

LITERATURE CITED

1. **Bestwick, R. K., C. A. Machida, E. Polonoff, and D. Kabat.** 1984. Frequent hereditable shutdown of murine retrovirus gene expression in murine cell lines. Mol. Cell. Biol. 4:908–914.
2. **Bird, A. P., and E. M. Southern.** 1978. Use of restriction enzymes to study eukaryotic DNA methylation. I. The methylation pattern in ribosomal DNA from *Xenopus laevis*. J. Mol. Biol. 118:27–48.
3. **Bondurant, M., R. Ramabhadran, M. Green, and W. S. M. Wold.** 1979. "sarc" Sequence transcription in moloney sarcoma virus-transformed nonproducer cell lines. J. Virol. 29:76–82.
4. **Caskey, C. T., and G. D. Kruh.** 1979. The HPRT locus. Cell 15:1–9.
5. **Chasin, L.** 1972. Nonlinkage of induced mutations in Chinese hamster cells. Nature (London) New Biol. 240:50–53.
6. **Chiswell, D. J., P. J. Enrietto, S. Evans, K. Quade, and J. A. Wyke.** 1982. Molecular mechanisms involved in morphological variation of avian sarcoma virus-infected rat cells. Virology 116:428–440.
7. **Copeland, N. G., K. W. Hutchison, and N. A. Jenkins.** 1983. Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. Cell 33:379–387.
8. **Feinstein, S. C., S. R. Ross, and K. R. Yamamoto.** 1982. Chromosomal position effects determine transcriptional potential of integrated mammary tumor virus DNA. J. Mol. Biol. 156:549–565.
9. **Gillespie, D. A. F., K. A. Hart, and J. A. Wyke.** 1985. Rearrangements of viral and cellular DNA are often associated with expression of Rous sarcoma virus in rat cells. Cell 41:279–287.
10. **Goff, S., P. Traktman, and D. Baltimore.** 1981. Isolation and properties of Moloney murine leukemia virus mutants: use of a rapid assay for release of virion reverse transcriptase. J. Virol. 38:239–248.
11. **Jaenisch, R., D. Jahner, P. Nobis, I. Simon, J. Lohler, K. Harbers, and D. Grotkopp.** 1981. Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell 24:519–529.
12. **Jahner, D., and R. Jaenisch.** 1980. Integration of Moloney leukemia virus into the germ line of mice: correlation between site of integration and virus activation. Nature (London) 287:456–458.
13. **Jolly, D. J., A. C. Esty, H. U. Bernard, and T. Friedmann.** 1982. Isolation of a genomic clone encoding human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. USA 79:5038–5041.
14. **Jolly, D. J., H. Okayama, P. Berg, A. C. Esty, D. Filpula, P. Bohlen, G. G. Johnson, J. E. Shively, T. Hunkapillar, and T. Friedmann.** 1983. Isolation and characterization of a full-length expressible cDNA for human hypoxanthine phosphoribosyltransferase. Proc. Natl. Acad. Sci. USA 80:477–481.
15. **Joyner, A., G. Keller, R. A. Phillips, and A. Bernstein.** 1983. Retrovirus transfer of a bacterial gene into mouse haematopoietic progenitor cells. Nature (London) 305:556–558.
16. **Joyner, A. L., and A. Bernstein.** 1983. Retrovirus transduction: segregation of the viral transforming function and the herpes simplex virus *tk* gene in infectious Friend spleen focus-forming virus thymidine kinase vectors. Mol. Cell. Biol. 3:2191–2202.
17. **Kawai, S., and H. Hanafusa.** 1973. Isolation of defective mutant of avian sarcoma virus. Proc. Natl. Acad. Sci. USA 70:3493–3497.
18. **Maniatis, T., A. Jeffrey, and D. G. Kleid.** 1975. Nucleotide sequence of the rightward operator of phage λ. Proc. Natl. Acad. Sci. USA 72:1184–1188.
19. **Mathey-Prevot, B., M. Shibuya, J. Samarut, and H. Hanafusa.** 1984. Revertants of partial transformants of rat fibroblasts infected with Fujinami sarcoma virus. J. Virol. 50:325–334.
20. **Miller A. D., R. J. Eckner, D. J. Jolly, T. Friedmann, and I. M. Verma.** 1984. Expression of a retrovirus encoding human HPRT in mice. Science 225:630–632.
21. **Miller, A. D., D. J. Jolly, T. Friedmann, and I. M. Verma.** 1983. A transmissible retrovirus expressing human hypoxanthine phosphoribosyltransferase (HPRT): gene transfer into cells obtained from humans deficient in HPRT. Proc. Natl. Acad. Sci. USA 80:4709–4713.
22. **Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara.** 1983. Inde-

Downloaded from mcb.asm.org by on September 14, 2009

pendent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell 32:1105–1113.
23. O'Rear, J. J., and H. M. Temin. 1982. Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, an avian retrovirus. Proc. Natl. Acad. Sci. USA 79:1230–1234.
24. Perkins, A. S., P. T. Kirschmeier, S. Gattoni-Celli, and I. B. Weinstein. 1983. Design of a retrovirus-derived vector for expression and transduction of exogenous genes in mammalian cells. Mol. Cell. Biol. 3:1123–1132.
25. Quintrell, N., S. H. Hughes, H. E. Varmus, and J. M. Bishop. 1980. Structure of viral DNA and RNA in mammalian cells infected with avian sarcoma virus. J. Mol. Biol. 143:363–393.
26. Searle, S., D. A. F. Gillespie, D. J. Chiswell, and J. Wyke. 1984. Analysis of the variations in proviral cytosine methylation that accompany transformation and morphological reversion in a line of Rous sarcoma virus-infected Rat-1 cells. Nucleic Acids Res. 12:5193–5210.
27. Sharp, J. D., N. E. Capecchi, and M. R. Capecchi. 1973. Altered enzymes in drug-resistant variants of mammalian tissue culture cells. Proc. Natl. Acad. Sci. USA 70:3145–3149.
28. Shimotohno, K., and H. M. Temin. 1981. Formation of infectious progeny virus after insertion of herpes simplex thymidine kinase gene into DNA of an avian retrovirus. Cell 26:67–77.
29. Shin, S. I. 1974. Nature of mutations conferring resistance to 8-AG in mouse cells. J. Cell Sci. 14:235–251.
30. Shtivelman, E., R. Zakut, and E. Canaani. 1984. Frequent generation of nonrescuable reorganized Moloney murine sarcoma viral genomes. Proc. Natl. Acad. Sci. USA 81:294–298.
31. Simon, D., H. Stuhlmann, D. Jahner, H. Wagner, E. Werner, and R. Jaenisch. 1983. Retrovirus genomes methylated by mammalian but not bacterial methylase are non-infectious. Nature (London) 304:275–277.
32. Sorge, J., D. Wright, V. D. Erdman, and A. E. Cutling. 1984. Amphotropic retrovirus vector system for human cell gene transfer. Mol. Cell. Biol. 4:1730–1737.
33. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.
34. Tabin, C. J., J. W. Hoffman, S. P. Goff, and R. A. Weinberg. 1982. Adaptation of a retrovirus as a eucaryotic vector transmitting the herpes simplex virus thymidine kinase gene. Mol. Cell. Biol. 2:426–436.
35. Temin, H. M. 1980. Origin of retroviruses from cellular moveable genetic elements. Cell 21:599–600.
36. Thilly, H., J. DeLuca, E. Furth, H. Hoppe, D. Kaden, J. Krolewski, H. Liber, T. Skopek, S. Slapikoff, R. Tizard, and B. Penman. 1980. Gene-locus mutation assays in diploid human lymphoblast lines, p. 331–364. *In* F. Deserres and A. Hollaender (ed.), Chemical mutagens, vol. 6. Plenum Publishing Corp., New York.
37. Thomson, L. H., and R. M. Baker. 1973. Isolation of mutants of cultured mammalian cells. Methods Cell Biol., p. 209–281.
38. Turner, D. R., A. A. Morley, M. Haliandoff, R. Kutlaca, and B. J. Sanderson. 1985. *In vivo* somatic mutations in human lymphocytes frequently result from major gene alterations. Nature (London) 315:342–345.
39. Varmus, H., and R. Swanstrom. 1982. Replication of retroviruses, p. 369–519. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. Varmus, H. E., N. Quintrell, and S. Ortiz. 1981. Retroviruses as mutagens: insertion and excision of a nontransforming provirus alter expression of a resident transforming provirus. Cell 25:23–36.
41. Varmus, H. E., N. Quintrell, and J. Wyke. 1981. Revertants of an ASV-transformed rat cell line have lost the complete provirus or sustained mutations in src. Virology 108:25–46.
42. Vogt, P. K. 1971. Spontaneous segregation of nontransforming viruses from cloned sarcoma viruses. Virology 46:939–946.
43. Watanabe, S., and H. M. Temin. 1983. Construction of a helper cell line for avian reticuloendotheliosis virus cloning vectors. Mol. Cell. Biol. 3:2241–2249.
44. Wei, C. M., M. Gibson, P. G. Spear, and E. M. Scolnick. 1981. Construction and isolation of a transmissible retrovirus containing the *src* gene of Harvey murine sarcoma virus and the thymidine kinase gene of herpes simplex virus type 1. J. Virol. 39:935–944.
45. Williams, D. A., I. R. Lemischka, D. G. Nathan, and R. C. Mulligan. 1984. Introduction of new genetic material into pluripotent haematopoietic stem cells of the mouse. Nature (London) 310:476–480.
46. Willis, R. C., D. J. Jolly, A. D. Miller, M. M. Plent, A. C. Esty, P. J. Anderson, H.-C. Chang, O. W. Jones, J. E. Seegmiller, and T. Friedmann. 1984. Partial phenotypic correction of human Lesch-Nyhan (hypoxanthine-guanine phosphoribosyltransferase-deficient) lymphoblasts with a transmissible retroviral vector. J. Biol. Chem. 259:7842–7849.
47. Zarling, D. A., and H. M. Temin. 1976. High spontaneous mutation rate of an avian sarcoma virus. J. Virol. 17:74–84.

Downloaded from mcb.asm.org by on September 14, 2009

TAB 11

Gene, 91 (1990) 179–184
Elsevier

This material may be protected by Copyright law (Title 17 U.S. Code)

GENE 03596

The methylation-free status of a housekeeping transgene is lost at high copy number

(Recombinant DNA; hypomethylation; *HTF* island; hydroxy-methylglutaryl CoA reductase; cholesterol)

Majid Mehtali[a,b], Marianne LeMeur[a] and Richard Lathe[a,c]

[a] *I.G.M.E.-CNRS and U184-INSERM, 11 rue Humann, 67085 Strasbourg Cedex (France) Tel. 88-37 12 55;* [b] *Transgene S.A. 11 rue de Molsheim, 67000 Strasbourg (France) Tel. 88 22 24 90;* [c] *AFRC Centre for Animal Genome Research, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ (U.K.) Tel. 31 668 4541*

Received by J.-P. Lecocq: 21 August 1989
Revised: 21 January 1990
Accepted: 22 February 1990

SUMMARY

Transgenic mouse lines were established bearing tandem arrays of a fusion construct comprising the promoter region of a housekeeping gene, *HMGCR*, encoding 3-hydroxy 3-methylglutaryl CoA reductase, linked to a bacterial *cat* reporter gene encoding chloramphenicol acetyltransferase (CAT). CAT activity was observed in all transgenic mouse tissues examined. The methylation state of the fusion transgene was determined. In non-transgenic mice the endogenous *HMGCR* promoter is devoid of methylation while flanking regions are extensively modified. In *HMGCR-cat* transgenic mice the fusion gene promoter was found to be similarly hypomethylated. However, the extent of hypomethylation varied with copy number: methylation-free status was progressively lost with increasing transgene copy number. Further transgenic mouse lines were constructed carrying a truncated *HMGCR* regulatory region linked to *cat*. Transgene expression and hypomethylation were observed in testis but not in any other tissue, and testis-specific methylation-free status was also lost at high copy number. Loss of hypomethylation at high copy number may indicate that saturable DNA-binding factors normally protect the *HMGCR* promoter from methylation.

INTRODUCTION

In contrast to the largely methylation-free genomes of bacteria and invertebrates, vertebrate genomes have extensive 5-methylation of C residues within CpG doublets (Grippo et al., 1968), and under-representation of CpG has been attributed to spontaneous deamination of 5-methyl-C (Bird, 1980). The relative insensitivity of vertebrate DNA to cleavage with enzymes whose recognition sequence contains a CpG doublet, for instance *Hpa*II (CCGG), is a consequence both of the rarity of CpG doublets and of cleavage inhibition by methylation of the recognition sequence. However, a small fraction of the genome is efficiently cleaved by *Hpa*II (Cooper et al., 1983) and such regions, '*Hpa*II-tiny-fragment' (HTF) islands, rich in unmethylated CpG doublets, appear to be associated with actively transcribed genes (Lavia et al., 1987) and particularly with the promoter regions of housekeeping genes (Bird, 1986; Gardiner-Garden and Frommer, 1987). Promoter regions of ubiquitously-expressed genes are unusual in that they are G/C-rich and lack consensus transcription initiation signals such as the 'TATA' or 'CAAT' boxes, and ubiquitous expression has been attributed to hypomethylation of control regions. (reviewed by Bird, 1986; 1987; Dynan, 1986; Gardiner-Garden and Frommer, 1987; Cedar, 1988). Methylation can correlate with diminished expression in vivo and in vitro (reviewed by Cedar, 1988);

Correspondence to: Dr. R. Lathe, CAGR, King's Buildings, West Mains Road, Edinburgh EH9 3JQ (U.K.) Tel. 31 668 4541; Fax 31 667 0164.

Abbreviations: bp, base pair(s); CAT, Cm acetyltransferase; *cat*, gene encoding CAT; Cm, chloramphenicol; HMGCR, 3-hydroxy-3-methylglutaryl CoA reductase; *HMGCR*, gene encoding HMGCR; *HTF*, *Hpa*II tiny fragment; kb, kilobase(s) or 1000 bp; nt, nucleotide(s).

0378-1119/90/$03.50 © 1990 Elsevier Science Publishers B.V. (Biomedical Division)

however, the causal link between hypomethylation and gene expression is unclear.

HMGCR (3-hydroxy-3-methylglutaryl CoA reductase) is a membrane-bound glycoprotein that catalyses a key step in the synthesis of cholesterol, an essential component of the cell membrane (Brown and Goldstein, 1980; Luskey, 1986). The *HMGCR* genes of hamster and human are characterized by a noncoding first exon and a long (ca. 3.5 kb) intervening sequence prior to the translation start site in the second exon (Reynolds et al., 1985; Luskey, 1987). We recently isolated the homologous mouse *HMGCR* gene and the promoter region, as for the hamster and human genes, bears the hallmarks of a *HTF* island in high G/C content and a CpG/GpC ratio close to unity (M.M. and R.L., unpublished).

We have used the transgenic mouse system to explore *HTF* island hypomethylation. Pronuclear microinjection of DNA into fertilized mouse eggs is now an established technique for generating mice bearing new gene combinations. In the majority of cases the injected DNA integrates at a single site as a multiple tandem repeat, and transgene copy number varies considerably between different transgenic lines (reviewed by Palmiter and Brinster, 1986). To explore possible variation in transgene functional status with copy number, two different fusion genes between the *HMGCR* promoter region and a *cat* reporter gene were constructed and introduced into the mouse germline. Employing these transgenic animals we endeavoured to study the relationship between transgene copy number, expression, and hypomethylation of the transgene promoter.

RESULTS AND DISCUSSION

(a) Transgenic animals bearing the *HMG1-cat* fusion construct

Construct *HMG1-cat* comprises a 5.5-kb *Bam*HI fragment of the mouse *HMGCR* promoter region, containing 1.35 kb of upstream sequence, the first (noncoding) exon and the first intron, linked to *cat* (Fig. 1A). The fusion construct was injected into fertilized mouse eggs and lines of transgenic mice were obtained carrying between 10 and 260 tandem copies of the transgene. These animals all express the fusion transgene in all tissues examined irrespective of transgenic line (Table I; also not presented): no proportionality was observed between expression level and copy number (M.M. and R.L., unpublished) as recorded (Palmiter and Brinster, 1986) in other transgenic systems.

(b) Methylation status of the *HMG1-cat* transgene

To assess the in vivo methylation status of the *HMGCR* promoter, we measured the extent to which template

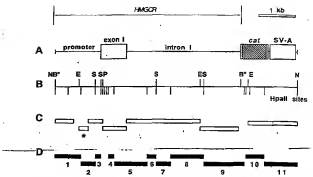


Fig. 1. Structure and analysis of the *HMGCR-cat* fusion transgene. (A) Structure of the injected construct *HMG1-cat*. The first *HMGCR* exon (noncoding) and the untranslated 5′ portion of the *HMGCR* second exon are indicated by open boxes; *cat*, hatched; the DNA sequence SV-A downstream from *cat* contains SX40 splice and polyadenylation signals derived from plasmid pSV2-CAT (Gorman et al., 1982). (B) Transgene restriction map. *Above*, restriction sites are B, *Bam*HI; B°, a *Bam*HI site destroyed during the cloning procedure; P, *Pvu*II, E, *Eco*RI, S, *Sma*I; flanking *Not*I sites (N) used for excision of the DNA construct prior to microinjection into fertilized mouse eggs are derived from the plasmid vector. *Below*, *Hpa*II/*Msp*I sites; the group of sites beneath the *Pvu*II site (P) is a cluster of seven sites over a region of 225 bp (spacing: 20, 25, 80, 30, 50 and 20 nt; unpublished data). (C) Hybridization probes employed to determine methylation status; the probe segment employed in the Southern analysis of Fig. 2 is marked with an asterisk. (D) DNA segments analysed for methylation inhibition of excision in Fig. 3.

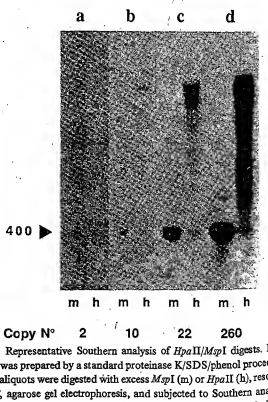


Fig. 2. Representative Southern analysis of *Hpa*II/*Msp*I digests. Liver DNA was prepared by a standard proteinase K/SDS/phenol procedure; 10 μg aliquots were digested with excess *Msp*I (m) or *Hpa*II (h), resolved by 1% agarose gel electrophoresis, and subjected to Southern analysis using a radiolabelled probe (marked with an asterisk in Fig. 1C) designed to detect the 400-bp segment 2 (arrowed) in Fig. 1D. Animals were: (a) wild-type; (b) transgenic line 74; (c) line 40; (d) line 78. The high-M_r bands of cleavage-resistant DNA are above the highest M_r marker employed (21.5 kb; not shown).

modification was able to inhibit excision of different transgene DNA segments from the genome. Total liver DNA from representative animals of different transgenic lines was cleaved either with *Hpa*II or with the methylation-insensitive isoschizomer *Msp*I, and after gel electrophoresis and Southern blotting examined for hybridization to separate probes (Fig. 1C) covering different regions of the *HMGCR* gene. The extent of methylation within each segment (Fig. 1D) of the transgene was assessed by the ratio of band intensities in the *Hpa*II and *Msp*I lanes. A typical experimental result is presented in Fig. 2 (zone analysed: segment 2 in Fig. 1D); data obtained for segments covering the entire transgene are compiled in Fig. 3.

Whereas excision of segments within the endogenous *HMGCR* promoter (two copies per diploid genome) in normal mice was not detectably blocked by methylation (<2%), flanking regions were resistant to digestion with *Hpa*II (Fig. 3a). In transgenic mice carrying ten copies of the *HMG1-cat* construct the *HMGCR* transgene promoter remained demethylated although the region devoid of methylation was narrower than that of the *HMGCR* gene in normal mice (Fig. 3b). However, methylation of the transgene promoter was observed to increase progressively in animals carrying 22 (35% methylation-inhibition of seg-

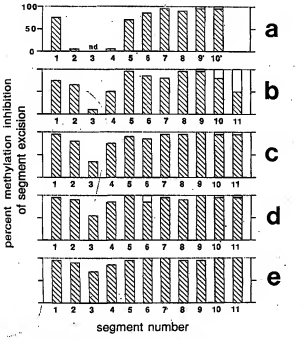


Fig. 3. Compilation of methylation data. Representative animals carrying different transgene copy numbers were: (a) wild type (2 copies of the *HMGCR* promoter); (b) transgenic line 74 carrying ten copies of the *HMG1-cat* fusion transgene); (c) line 40, 22 copies; (d) line 80, 45 copies; (e) line 78, 260 copies. Vertical axis: the extent of methylation inhibition of segment excision, determined as the ratio of the intensities of the relevant bands in the *Hpa*II and *Msp*I lanes of liver DNA (e.g. Fig. 2). Unshaded areas above certain histogram bars indicate 'greater-than' values. Horizontal axis: segment number (see Fig. 1D); internally-labelled probes used to determine methylation status (Mehtali, 1988) are presented in Fig. 1C. Quantitative scanning densitometry of autoradiograms (GS300 Scanner, Hoeffer Scientific) was used to determine the *Hpa*II/*Msp*I excision ratio. Segment 1 comprises two equally-sized *Hpa*II fragments that were not resolved by gel electrophoresis; segment 10 overlaps an *Hpa*II 'slow' site within *cat* that is partially refractory to *Hpa*II cleavage. Segments 9 and 10 represent the next two *Hpa*II fragments present in the endogenous *HMGCR* gene but not in the *HMG1-cat* transgene. For technical reasons the excision ratio for segment a3 was not determined (*nd*).

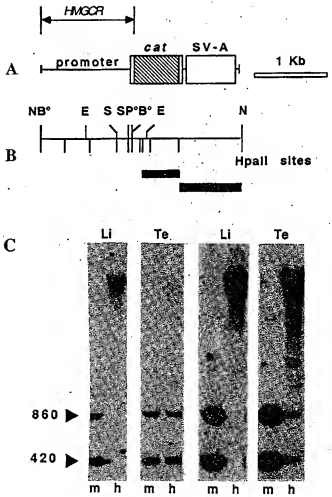


Fig. 4. Methylation state of the *HMG2-cat* transgene. (A) Structure of *HMG2-cat*: this construct is a derivative of *HMG1-cat*, in which the region between the *Pvu*II site in the *HMGCR* first exon and the site marked B° at the 5' end of *cat* in *HMG1-cat* (see Fig. 1A,B) has been removed. (B) Restriction site nomenclature is as in Fig. 1B; the solid bars given below indicate the segments detected by the hybridization probe employed (DNA fragment comprising *cat* and the SV40 splice and polyadenylation signals, not shown). (Panels C) Southern hybridization of *Msp*I (m) and *Hpa*II (h) digests of total liver (Li) and testis (Te) DNA from representative animals bearing eleven (line 61) or 170 copies (line 76) of the fusion transgene. *Hpa*II fragments detected (sizes in bp) are arrowed. High-M_r band sizes and all experimental conditions are as in Fig. 2 legend.

TABLE I

CAT activity in tissues of *HMGCR-cat* transgenic mice

Construct[a]	Line	Copy No.[b]	CAT activity[d] in mouse tissues[c]						
			ta	li	in	ki	br	st	te
HMG1-cat	74[e]	10	12	5	21	5	62	31	59
HMG1-cat	40[e]	22	17	20	107	5	90	72	na
HMG1-cat	40	22	30	90	95	15	80	110	140
HMG1-cat	80[e]	45	6	32	21	7	70	23	na
HMG1-cat	78[e]	260	0.6	0.5	2	0.6	0.6	0.5	32
HMG2-cat	61[f]	11	0.6	0.9	0.7	0.6	0.7	0.8	135
HMG2-cat	76[f]	170	0.5	1.6	0.7	0.6	0.7	0.8	208

[a] Construct p*HMG1-cat* contains a 5.5-kb *Bam*HI segment of the mouse *HMGCR* gene (Mehtali, 1988) comprising the mouse *HMGCR* promoter region, the first (untranslated) exon, and the first intron, linked to the bacterial *cat* gene (Fig. 1A) and propagated in plasmid pPolyIII-I (Lathe et al., 1987). The downstream *Bam*HI site used in this construction lies immediately beyond the *HMGCR* exon II splice acceptor site (Mehtali, 1988; Gautier et al., 1989; M.M. and R.L., in preparation). In *pHMG2-cat* the region between the *Pvu*II site in the *HMGCR* first exon and the *BamHI*° site at the 5′ end of *cat* (see Fig. 1A) was removed: the structure of the fusion gene is presented in Fig. 4. Plasmid DNAs were propagated on *E. coli* 1106 (*ihr leu thi hsdS*). *HMG-cat* fusion genes were excised from the vector by *Not*I digestion and purified by sucrose gradient centrifugation (10–30% sucrose, 35 000 rpm, Beckman SW41 rotor, 14 h, 20°C). Two hundred copies were injected into fertilized eggs (C57B1/6 × SJLF1 hybrids) and the presence of the transgene was detected by Southern blot analysis of DNA from the tails of four-week-old animals (Palmiter et al., 1982). Transgenic lines were established by systematic back-crossing with C57B1/6 × SJLF1 hybrids.

[b] Copy numbers were determined by quantitative densitometry of Southern blots; all transgenic animals carry the transgene as an unrearranged tandem repeat integrated at a single autosomal location (data not presented).

[c] Tissues were: ta, tail; li, liver; in, intestine; ki, kidney; br, brain; st, stomach; te, testis.

[d] CAT activity in tissue homogenates (ultra-turrax, in 0.25 M Tris-HCl pH 7.8, 1 mM phenylmethylsulfonylfluoride) was determined, after heat treatment (10 min, 65°C) and clarification, by the transfer of radioactivity from ^{14}C acetyl CoA to ethyl-acetate-soluble Cm in a standard assay (Sleigh, 1986) at a constant protein concentration as measured using a commercial assay kit (BioRad), and is expressed as the incorporation of ^{14}C acetyl into Cm (cpm × 10^3) per 100 μg protein. The background value for CAT activity in negative extracts was 0.5. na, not applicable.

[e] Animals used for analysis in Fig. 2 (founder transgenic male of line 78; founder transgenic females of lines 40 and 80, and a second generation male for line 74; the remaining line 40 animal was a second generation male; M.M. and R.L., in preparation).

[f] Animals used for analysis were second generation males.

ment excision, Fig. 3c), 45 (48%) or 260 copies (78%) (Fig. 3d,e). The methylation pattern appeared identical in other tissues examined (intestine, testis) though the extent of the hypomethylated region was slightly larger in testis than in the other tissues (data not presented). The strength of the hybridization signal originating from the transgene unfortunately precluded examination of the methylation pattern of the endogenous *HMGCR* gene in these transgenic mice.

(c) *HMG2-cat* transgenic mice

To determine whether there is a general correlation between increased transgene copy number and loss of methylation-free status we constructed a deletion derivative of the *HMG1-cat* construct, *HMG2-cat*, in which a sub-fragment of the *HMGCR* promoter region is linked directly to *cat* (Fig. 4A). *HMG2-cat* was introduced into the mouse germline and tissues from transgenic animals were analysed for CAT activity (Table I). In the two lines examined (61 and 67; Table I) activity was detected in testis but not in other tissues. The methylation state of the fusion transgene in lines 61 and 76 was examined by *Hpa*II or *Msp*I digestion and Southern hybridization to probes designed to detect the reporter gene segment of the transgene (Fig. 4B) or the *HMGCR* promoter region (data not presented).

In liver (Fig. 4C) and intestine (data not presented), tissues in which no expression is observed (Table I), the fusion transgene was extensively methylated in both transgenic lines. In testis of mice harboring eleven copies of the transgene (line 61), a tissue in which the transgene is expressed, the *HMG2-cat* promoter and surrounding regions were essentially devoid of methylation (Fig. 4C; also data not presented). In contrast, the transgene in testis of line 76 (harboring 170 copies) was found to be substantially methylated (Fig. 4C), supporting a link between increased copy number and loss of methylation-free status. Nevertheless, in testis of line 76 a small proportion of transgene copies appeared to be devoid of methylation; these few unmethylated copies may be responsible for the observed transgene expression level. Indistinguishable methylation patterns were obtained in all cases irrespective of whether the probe employed covered the reporter gene segment (Fig. 4C) or the *HMGCR* promoter region (data not presented). It is of note that hypomethylation of the

HMGCR promoter region in testis of line 61 now extends into the adjacent bacterial *cat* gene (Fig. 4). Similar hypomethylation of *cat* was observed in testis of a further transgenic mouse line (No 60) harboring 15 copies of the *HMG2-cat* transgene (data not presented).

(d) Conclusions

HTF islands are associated with control regions of active genes, particularly the housekeeping genes (Bird, 1986; Lavia et al., 1987; Gardiner-Garden and Frommer, 1987). It was previously reported that the methylation-free *HTF* island of the *Thy-1* gene is maintained when the intact gene is introduced into the mouse germline (Kolsto et al., 1986). We describe here that hypomethylation of a housekeeping gene (*HMGCR*) promoter is dependent upon transgene copy number, and methylation-free status of the *HTF* island at the 5′ end of a *HMGCR-cat* fusion transgene, *HMG1-cat*, is lost at increasing copy-number. In further transgenic mice bearing the *HMG2-cat* deletion derivative of the fusion construct, CAT activity and transgene hypomethylation were only detected in testis. It is of note that ectopic expression of tissue-specific transgenes in testis has been observed previously (Lacy et al., 1983; Shani, 1986; Al-Shawi et al., 1988; and our unpublished data). As observed in *HMGCR-cat* transgenic animals, methylation-free status in testis of *HMG2-cat* transgenic animals was also lost at high copy number.

One possible explanation for loss of hypomethylation at high copy number is out-titration of regulatory proteins or other binding factors that protect the DNA from methylation. Our data do not exclude the possibility that out-titration of a demethylase activity (e.g., Razin et al., 1986) might also reduce methylation level. However, methylation of the transgene promoter at high copy number demonstrates that *HTF* island DNA is not intrinsically resistant, in vivo, to methylation of C residues within CpG doublets. Because saturable factors thus appear to protect DNA from methylation in vivo, factor binding seems likely to precede methylation. In consequence, it would appear unlikely that the methylation status of the *HMGCR* promoter itself plays a major role in determining the extent of factor binding in vivo.

Despite the fact that the *HMGCR* promoter present on the transgene can outnumber the endogenous *HMGCR* promoter by a factor of 100, we have not detected any alteration in the expression of the endogenous gene (M.M., R.L. and G. Boukamel, unpublished data), in agreement with the conclusions of a study (Davis and MacDonald, 1988) using a rat elastase I transgene. Because high transgene copy number appears to be without effect on the expression of the endogenous *HMGCR* gene, binding factors may be only locally out-titrated, possibly arguing for restricted diffusion of binding factors (see Richetti et al., 1988).

Although we report saturable hypomethylation of the *HMGCR* promoter, this result may contrast with the report of Kolsto et al. (1986) who describe hypomethylation of a hybrid *Thy-1* gene promoter in mice bearing 60 copies of the transgene. However, the presence of extraneous sequences in *Thy-1* transgenic mice (Grosveld and Kollias, 1988) complicates interpretation. It therefore remains unclear whether loss of methylation-free status and/or local out-titration of binding factors is a general feature of large transgene arrays.

We also report that the hypomethylated region of the *HMGCR-HTF* island can extend into adjacent bacterial DNA (*cat*). This phenomenon was only observed when the *cat* reporter gene was linked directly to the *HMGCR* promoter region (construct *HMG2-cat*) and not when the *cat* segment was separated from the *HMGCR* regulatory region by 4 kb of intervening *HMGCR* genomic DNA (construct *HMG1-cat*). Hypomethylation of adjacent bacterial DNA was only observed in testis, the only tissue in which transgene expression was detected, and we speculate that cooperative binding of factors (e.g., Phillips et al., 1989; see also Murray and Grosveld, 1987) to the *HMGCR* promoter and to adjacent CpG-rich bacterial DNA may be responsible. We cannot however exclude the possibility that transcription per se can contribute to under-methylation.

Taken together, our data argue that the cytosine-methylase passively methylates DNA according to its accessibility/affinity for the methylase. Because DNA methylation can inhibit gene activity, passive methylation of transcriptionally inactive regions may contribute to the repressed state. Lower eukaryotes and invertebrates lack detectable DNA methylation, and the large genome sizes of vertebrates and plants may provide a selective advantage for DNA methylation (Antequera and Bird, 1988), for instance by marking untranscribed and passively-methylated DNA for higher-order condensation.

ACKNOWLEDGEMENTS

We would like to thank C. Marfing for skillful microinjection and P. Meyer and I. Kuhn for technical assistance. C. Tomasetto, J.O. Bishop, A.J. Clark, A. Pavirani, P. Gerlinger and P. Chambon are thanked for helpful discussions. This work was supported by grants from the CNRS, INSERM, Centre de Lutte contre le Cancer, ARC and the FRMF.

REFERENCES

Al-Shawi, R., Burke, J., Jones, C.T., Simons, J.P. and Bishop, J.O.: A MUP promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. Mol. Cell. Biol. 8 (1988) 4821–4828.

Antequera, F. and Bird, A.P.: Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J. 7 (1988) 2295–2299.

Bird, A.P.: DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8 (1980) 1499–1504.

Bird, A.P.: CpG islands and the function of DNA methylation. Nature 321 (1986) 209–213.

Bird, A.P.: CpG islands as gene markers in the vertebrate nucleus. Trends Genet. 3 (1987) 342–347.

Brown, M.S. and Goldstein, J.L.: Multivalent feedback regulation of HMGCoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth. J. Lipid Res. 21 (1980) 505–517.

Cedar, H.: DNA methylation and gene activity. Cell 53 (1988) 3–4.

Cooper, D.N., Taggart, M.H. and Bird, A.P.: Unmethylated domains in vertebrate DNA. Nucleic Acids Res. 11 (1983) 647–658.

Davis, B.P. and MacDonald, R.J.: Limited transcription of rat elastase I transgene repeats in transgenic mice. Genes Dev. 2 (1988) 13–22.

Dynan, W.S.: Promoters for housekeeping genes. Trends Genet. 2 (1986) 196–197.

Gardiner-Garden, M. and Frommer, M.: CpG islands in vertebrate genomes. J. Mol. Biol. 196 (1987) 261–282.

Gautier, C., Mehtali, M. and Lathe, R.: A ubiquitous mammalian expression vector, pHMG, based on a housekeeping gene promoter. Nucleic Acids Res. 17 (1989) 8389.

Gorman, C.M., Moffat, L.F. and Howard, B.H.: Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell. Biol. 2 (1982) 1044–1051.

Grippo, P., Daccararino, M., Parrasie, E. and Scarano, E.: Methylation of DNA in developing sea urchin embryos. J. Mol. Biol. 36 (1968) 195–208.

Grosveld, F. and Kollias, G.: Ectopic expression of Thy-1 in the kidneys of transgenic mice induces functional and proliferative abnormalities. Cell 54 (1988) 920.

Kolsto, A.-B., Kollias, G., Giguere, V., Isobe, K.-I., Prydz, H. and Grosveld, F.: The maintenance of methylation-free islands in transgenic mice. Nucleic Acids Res. 14 (1986) 9667–9678.

Lacy, E., Roberts, S., Evans, E.P., Burtenshaw, M.D. and Costantini, F.D.: Foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and in inappropriate tissues. Cell 34 (1983) 353–358.

Lathe, R., Vilotte, J.L. and Clark, A.J.: Plasmid and bacteriophage vectors for excision of intact inserts. Gene 57 (1987) 193–201.

Lavia, P., MacLeod, D. and Bird, A.: Coincident start sites for divergent transcripts at a randomly selected CpG rich island of mouse. EMBO J. 6 (1987) 2773–2779.

Luskey, K.L.: Structure and expression of 3-hydroxy 3-methylglutaryl coenzyme A reductase. Ann. N.Y. Acad. Sci. 478 (1986) 249–254.

Luskey, K.L.: Conservation of promoter sequence but not complex intron-splicing pattern in human and hamster genes for 3-hydroxy 3-methylglutaryl coenzyme A reductase. Mol. Cell. Biol. 7 (1987) 1881–1893.

Mehtali, M.: Ph.D. thesis, University of Strasbourg, 1988.

Murray, E.J. and Grosveld, F.: Site-specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation-mediated suppression. EMBO J. 6 (1987) 2329–2335.

Palmiter, R.D. and Brinster, R.L.: Germ line transformation of mice. Annu. Rev. Genet. 20 (1986) 465–499.

Palmiter, R.D., Chen, H.Y. and Brinster, R.L.: Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell 29 (1982) 701–710.

Phillips, S.E.V., Manfield, I., Parsons, I., Davidson, B.E., Rafferty, J.B., Somers, W.S., Margarita, D., Cohen, G.N., Saint-Girons, I. and Stockley, P.G.: Cooperative tandem binding of *met* repressor of *Escherichia coli*. Nature 341 (1989) 711–715.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D. and Cantoni, G.L.: Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA methylation during differentiation. Proc. Natl. Acad. Sci. USA 83 (1986) 2827–2831.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G. and Luskey, K.L.: HMG CoA reductase: a negatively-regulated gene with unusual promoter and 5' untranslated regions. Mol. Cell. Biol. 7 (1985) 1881–1893.

Richetti, M., Metzger, W. and Heumann, H.: One-dimensional diffusion of *Escherichia coli* DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. Proc. Natl. Acad. Sci. USA 85 (1988) 4610–4614.

Shani, M.: Tissue-specific and developmentally-regulated expression of a chimeric actin-globin gene in transgenic mice. Mol. Cell. Biol. 6 (1986) 2624–2631.

Sleigh, M.: A non-chromatographic assay for expression of the chloramphenicol-acetyl-transferase gene in eucaryotic cells. Anal. Biochem. 156 (1986) 257–262.

TAB 12

Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 1330–1334, February 1991
Cell Biology

Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes

(gene therapy/retroviral vectors/adenosine deaminase/neomycin phosphotransferase/severe combined immunodeficiency)

THEO D. PALMER*, GUY J. ROSMAN*, WILLIAM R. A. OSBORNE†, AND A. DUSTY MILLER*

*Program in Molecular Medicine, Fred Hutchinson Cancer Research Center, Seattle, WA 98104; and †Department of Pediatrics, University of Washington, Seattle, WA 98195

Communicated by E. Donnall Thomas, November 14, 1990

ABSTRACT **Genetically engineered fibroblasts have been successfully used to produce therapeutic proteins in animals, but sustained production of the proteins has not been achieved. This limits the potential of fibroblast-mediated gene therapy in humans. We have studied the phenomenon of decreased production in rats by using retroviral vectors carrying genes encoding human adenosine deaminase and neomycin phosphotransferase. While transplanted skin fibroblasts containing vector sequences persisted at constant levels for at least 8.5 mo, vector expression decreased by >1500-fold after 1 mo. Cellular or antibody-mediated immune responses were not detected in transplanted animals, and expression could not be restored in fibroblasts recultivated from the grafts. This phenomenon is reminiscent of sequence-specific gene inactivation observed in other cell types. Because genetic manipulation and expression of foreign proteins did not affect survival of the transplanted cells, effective long-term therapy may be possible with the use of alternative gene regulatory elements.**

Although many somatic cell types are potential gene therapy targets for treatment of genetic or acquired disease (1), skin fibroblasts are attractive because they are easy to obtain and transplant and can be rapidly grown to large numbers in culture. Immortalized fibroblasts make a variety of secretory products after genetic modification and transplantation into animals (2–7), but these are ultimately not appropriate models for gene therapy because these cell lines grow uncontrollably and often form tumors in recipient animals. Primary fibroblasts, however, could be used for human therapy. In animals, primary embryo or skin fibroblasts produce clotting factor IX at systemic levels that approach therapeutic utility (5, 6), but production of the protein for >1 mo has not been achieved. The gradual decrease in protein levels might be caused by immune response against the foreign protein, poor survival of the transplanted cells, or inactivation of the transferred genes.

To address these issues, we have studied the transfer of a human adenosine deaminase (hADA)-encoding gene into skin fibroblasts of inbred rats. By using hADA in place of factor IX, we could avoid several complicating problems associated with production of a foreign secreted protein: hADA should not be immunogenic because it is an intracellular protein; hADA is localized at the site of the graft, and detection is not dependent on systemic distribution; and finally, sensitive assays make even low-level hADA readily detectable above endogenous rat adenosine deaminase (ADA). By careful examination of transplanted tissues, we show here that expression of hADA had no effect on cell survival, but that transplanted cells gradually inactivate retrovirally transferred genes.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

MATERIALS AND METHODS

Primary Cell Culture. Primary skin fibroblasts were isolated by standard methods (8) from either human foreskin or forearm biopsies, or from inbred Fischer 344 rats. Cells were grown at 37°C in a 10% CO_2 atmosphere with Dulbecco's modified Eagle's medium (DMEM) containing 5 g of glucose per liter, 10% (human cells) or 15% (rat cells) fetal bovine serum, and amphotericin B, penicillin, and streptomycin antibiotics. To avoid potential complications of *in vitro* aging, human and rat cell cultures were used soon after establishment of the primary culture. In general, human fibroblasts were used within 3 to 20 population doublings of the primary culture, and rat fibroblasts were used within 3 to 5 population doublings.

Vectors. Except for pLASΔN, all vectors have been described (9). To construct pLASΔN, a deletion was introduced into the gene for neomycin phosphotransferase (neo) of pLASN by removing 400 base pairs (bp) between *Fsp* I and *Rsr* I. High-titer helper-free virus stocks were produced with PA317 amphotropic packaging cells (10), as described (11). Virus titers measured on NIH 3T3 cells ranged from 10^6 to 8 × 10^6 colony-forming units (cfu) per ml. As noted in previous studies, the titers were generally 10-fold lower on normal diploid fibroblast strains from either humans or rats (6, 9, 12, 13).

Transplantation. All fibroblast transplants were performed between inbred Fischer 344 rats cared for in accordance with institutional guidelines. Skin fibroblasts were isolated, cultured, and cast in collagen matrices as described (6). For dermal equivalent transplants, the matrix was placed in a circular full-thickness skin wound and protected with a burn dressing consisting of a collagen/condroitin sulfate matrix bound to a silicone backing (Integra, Marion Merrell Dow) similar to that described by Yannas *et al.* (14). The dressing was glued at the edges to the surrounding skin with cyanoacrylate glue (Nexaband Liquid, CRX Medical, Raleigh, NC). The graft was covered with nonadherent dressing, and the rat was wrapped with an elastic bandage. Subcutaneous ampicillin was given 1 day before transplantation and continued for 10 days.

Polymerase Chain Reaction (PCR). Genomic DNA (250 ng) was amplified in a total reaction volume of 55 μl (0.01% gelatin/50 mM KCl/10 mM Tris, pH 8.5/2 mM $MgCl_2$/0.1 mM each of dNTP/0.1 μM [α-^{32}P]dTTP or [α-^{32}P]dCTP/1.5 units of *Taq* polymerase/125 ng each of Neo1 and Neo5 primers). Primer sequences were as follows; Neo1, 5′-CAAGATGGATTGCACGCAGG-3′; and Neo5, 5′-CCCGC-TCAGAAGAACTCGTC-3′, respectively. In addition, 5 fg of plasmid pLASΔN (Fig. 1) was added to all samples as an

Abbreviations: neo, neomycin phosphotransferase; ADA, adenosine deaminase; hADA, human ADA; LTR, long terminal repeat; SV40, simian virus 40; cfu, colony-forming units; PNP, purine nucleoside phosphorylase; MoMLV, Moloney murine leukemia virus; PCR, polymerase chain reaction.

internal control. To guarantee that each sample received an equal amount of pLASΔN, an aliquot of plasmid was added to a complete reaction mixture, and then the mixture was divided equally between DNA samples. Amplification consisted of 25–30 cycles of 95°C for 2 min followed by 72°C for 5 min. The reaction products were resolved in 5% nondenaturing acrylamide gels, which were then dried onto filter paper and exposed to film at −70°C. Two PCR products were expected, a full-length neo band (*neo*) from the integrated LASN vector and a shorter band (*Δneo*) from pLASΔN. Band intensities (arbitrary units) were determined by computer analysis of a digitized video image (Biosystems image analysis system). Briefly, bands were defined as inclusive groups of pixels with optical densities above background. An individual band intensity is the sum optical density for all pixels within a band. To avoid problems with film saturation or reciprocity failure, several exposures of each gel were analyzed. The intensity of the *Δneo* sequence amplified from pLASΔN was used to correct for differences in amplification efficiency between samples (see Fig. 4).

Cytotoxic Lymphocyte Assays. Lymphocytes from proximal lymph nodes were used to lyse ^{51}Cr-labeled fibroblasts. Target fibroblasts were loaded with ^{51}Cr by mixing an equal volume of ^{51}Cr (5 mCi per ml in saline; 1 Ci = 37 GBq) with a single-cell suspension of fibroblasts (10^7 cells per ml) in DMEM/20% fetal bovine serum. The mixture was incubated for 1 hr at 37°C. The cells were then washed three times and resuspended in RPMI 1640/10% heat-inactivated fetal bovine serum. Specific cell lysis was measured by mixing 4×10^3 fibroblasts with different numbers of lymphocytes in V-bottom 96-well dishes. The cell mixtures were pelleted and incubated at 37°C for 4 hr. Spontaneous release of ^{51}Cr was measured on fibroblasts alone, and 1% Triton X-100 lysates were used to measure maximal release. Specific lysis was calculated as follows: % specific lysis = (experimental cpm − spontaneous cpm)/(maximum cpm − spontaneous cpm) × 100.

RESULTS

Expression of hADA in Cultured Fibroblasts. Several retroviral vectors encoding hADA (Fig. 1) were tested for hADA expression in primary diploid fibroblasts from humans and rats and in NIH 3T3 immortalized mouse fibroblasts (Table 1). Primary isolates of skin fibroblasts or secondary passages of a primary isolate were infected with limiting amounts of virus to ensure single-copy integration and were then treated with G418. Bulk populations of G418-resistant cells were then assayed for hADA. All three retroviral vectors produced significant amounts of human ADA, but LASN and LNCA were most efficient in human fibroblasts.

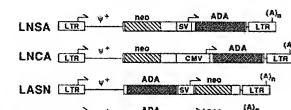


FIG. 1. ADA vectors. Retroviral vectors LNSA, LNCA, and LASN use three different promoters to express ADA; the simian virus 40 (SV40) early promoter (SV), the cytomegalovirus immediate early promoter (CMV), and the Moloney murine leukemia virus (MoMLV) promoter in the long terminal repeat (LTR). pLASΔN was used as a plasmid control for PCR reactions. Arrows, transcription initiation sites; $(A)_n$, polyadenylylation sites; ψ^+, packaging signal.

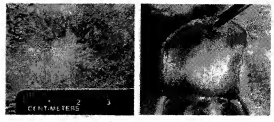


FIG. 2. Outer (*A*) and inner (*B*) surfaces of a dermal equivalent tissue at 8.5 mo. (Magnification is the same in both images.)

Enzyme activities in human fibroblasts infected with LNCA vector were so high that only ≈5×10^7 cells (a number feasible to transplant) would provide enough ADA to correct the biochemical defect in humans suffering from ADA deficiency (see ref. 13 for method of estimation). To examine long-term production of hADA in an animal model, the LASN vector, which was the most active vector in rat fibroblasts, was used to introduce hADA into primary skin fibroblasts of inbred Fischer 344 rats.

hADA Production and Cell Survival After Transplantation. Genetically modified primary rat skin fibroblasts were cast in collagen matrices and used as dermal equivalent grafts. The grafts were easily recognizable for over 8 mo, having a hairless upper surface (Fig. 2*A*) and a vascularized but pale lower surface (Fig. 2*B*). The grafted tissues were removed at various times after transplantation and cut into pieces for ADA and vector DNA analysis, and for reinitiation of fibroblast cultures. The pattern of hADA production was similar to that reported for factor IX (6); while hADA was clearly detected early, activities dropped to undetectable levels at ≈1 mo (Fig. 3, Table 2). From the limit of detection for hADA (≈1% of endogenous rat ADA) and the percentage of donor cells in the tissue sample at 1 mo (30%, Table 2), we estimate the drop from 50-fold above endogenous ADA levels (Table 1) to <1% of endogenous levels represents >1500-fold reduction in hADA levels.

The decrease in hADA levels directly contrasts with the continued presence of vector DNA in the transplanted tissue. PCR analysis of DNA extracted from tissues shows vector sequences to persist at undiminished levels for at least 8.5 mo (Fig. 4, Table 2). Although PCR analysis is inherently variable, appropriate internal controls and replicate assays (Fig. 4) allow rough quantitation of vector sequences. For exam-

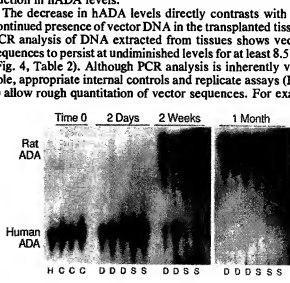


FIG. 3. Human and rat ADA in transplanted tissues. Equal amounts of total ADA were analyzed by starch gel electrophoresis as described (16). Although not shown, loading of equal amounts of total protein showed rat ADA to remain relatively constant in rat cell samples. Lanes: H, human T-cell sample; C, cultured cells before transplantation; D, dermal equivalent transplants; S, subcutaneous transplants of fibroblast-containing collagen matrices. The subcutaneous transplants behaved like the dermal equivalent transplants but were much more difficult to localize at long times after transplantation and are not further discussed.

Table 1. ADA activity in cultured fibroblasts

Cells	ADA activity, μmol per hr per mg of protein			
	No vector	LNSA vector	LNCA vector	LASN vector
Normal human fibroblasts	0.9 ± 0.4	8.2 ± 2.1	120 ± 55	80 ± 45
ADA⁻ human fibroblasts	<0.01	6.2 ± 2.6	120 ± 20	80 ± 40
Rat fibroblasts	0.9 ± 0.1	6.6 ± 2.3	4.9 ± 0.2	50 ± 25
NIH 3T3 Cells	2.2 ± 0.2	4.8 ± 1.3	5.9 ± 1.8	9.5 ± 4.7

Cells were infected with the indicated vectors and selected for G418 resistance. ADA was determined as described (15) for polyclonal (>10^5) populations, and the values given are means ± SDs for 2–12 separate primary strains.

ple, three transplants harvested at 8.5 mo had mean ± SD values (n = 3 or 4) of 61 ± 40, 63 ± 33, and 55 ± 28% vector-positive cells. This result yields an average value of 60% vector-positive cells at 8.5 mo (Table 2). Clearly, many cells contain vector sequences after 8.5 mo. Thus vector-infected cells persisted long-term, but expression of the transferred hADA gene was severely decreased.

Immune Responses to Transplanted Fibroblasts. Immune responses against vector-expressing fibroblasts might have selected for rare nonexpressing cells. To test this, five rats received dermal equivalent transplants of LASN-expressing cells. One month later the grafts were removed and shown to contain but not express vector sequences. To boost potential immune responses, the animals received s.c. injections of 10^7 additional cells when the grafts were removed. Ten days later lymphocytes from lymph nodes, spleens, and peripheral blood were assayed for specific lysis of LASN-expressing target fibroblasts. Plasmid from the rats was also tested for antibodies against the LASN-expressing fibroblasts by adding plasma and purified complement to the fibroblasts and scoring cell lysis by trypan blue dye exclusion. There was no increase in antibody-mediated cell lysis after transplantation (data not shown) or in cell lysis by lymphocytes from transplanted animals compared with that obtained by using nontransplanted animals (Fig. 5). In contrast, control animals injected with human fibroblasts generated strong cytolytic responses against the human cells.

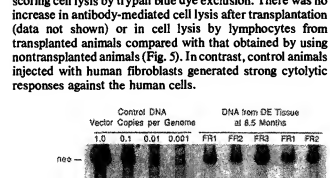


FIG. 4. PCR analysis of transplanted cells. PCR was used rather than Southern analysis because of the limited amount of DNA recovered from each tissue sample. (*Left*) Amplified products from control DNA containing known amounts of integrated LASN vector. (*Right*) Amplified products from three dermal equivalent (DE) tissues explanted at 8.5 mo (replicate assays are shown for transplants FR1 and FR2). Each sample contains the same amount (5 fg) to pLASΔN plasmid, which produces a shorter *neo*-specific band (Δ*neo*) than produced from integrated vector sequences (*neo*). Ratios of *neo*/Δ*neo* band intensities for control DNAs were plotted against vector copies per genome on logarithmic graph paper; an estimated best-fit line was drawn through plotted points. Control and experimental *neo*/Δ*neo* values were then converted to % vector containing cells by using this plot. Although background in this assay falls between 1% and 0.1%, sensitivity can be enhanced by increasing number of amplification cycles.

Although strong immune responses were not detected, low-level selection against cells expressing hADA or *neo* could have caused the gradual decrease in expression. To determine whether vector expression inhibited the survival of transplanted cells, naive or vector-expressing fibroblasts from female rats were transplanted onto male hosts. At 2 mo after transplant, cultures were initiated from the grafted tissues, and metaphase cells from each culture were examined for female or male karyotype (17). The average fraction of female cells (XX karyotypes/total karyotypes) from duplicate transplants were 0.41 ± 0.10 for naive cells and 0.34 ± 0.06 for vector-expressing cells. Twenty-three to 41 metaphase figures were karyotyped from each culture, and the values given are the means ± SDs for two transplants of each cell type. This result shows that genetic manipulation and expression of vector-encoded proteins has no significant effect on the survival of the transplanted fibroblasts and suggests that vector down-regulation was the major cause of decreased hADA expression.

Vector Expression in Cells Cultured from Skin Grafts. Fibroblasts were easily cultured from the dermal equivalent grafts and were found positive for vector DNA, although at somewhat lower levels than seen in the graft (Table 2). The presence of vector sequences in culture shows that vector-containing cells in the grafts were viable. Expression of hADA in the cultured fibroblasts decreased as a function of time that the fibroblasts had been in animals (Table 2), similar to the decrease seen in grafts. We also measured ability of the cultured fibroblasts to form colonies in G418 as a measure of SV40 promoter-directed *neo* gene expression (Table 2). Cultures initiated from grafts that were removed soon after transplantation showed a high proportion of G418-resistant cells, whereas those removed at 8.5 mo after transplantation showed no G418-resistant cells (<0.001%). These results show that decreased expression was not from the quiescent (i.e., nonreplicative) state of the fibroblasts *in vivo* because

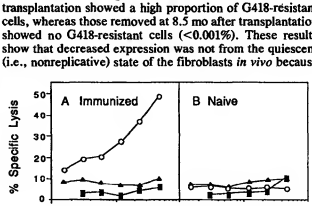


FIG. 5. Cytotoxic T lymphocyte assays. Lymph node lymphocytes from transplanted animals were mixed with ^{51}Cr-loaded fibroblasts, and ^{51}Cr release was used to measure specific cytolytic activity. (*A*) Lymphocytes were harvested from animals transplanted with target cells. (*B*) Lytic activity of lymphocytes from nontransplanted animals on same targets. ○, Human fibroblasts; ■, LASN-infected rat fibroblasts; ▲, uninfected rat fibroblasts.

Table 2. ADA vector persistence and expression in transplanted rat fibroblasts

Cells transplanted	Time	Transplanted tissue		Cells cultured from tissue		
		Vector positive, %	ADA/PNP	Vector positive, %	ADA/PNP	G418^r colonies, %
LASN infected	0			100	3.6	60
	2 days	40	1.9	30	3.2	46
	2 weeks	30	0.34	13	0.87	8
	1 mo	30	0.23	9	0.13	0.03
	8.5 mo	60	0.13	6	0.14	<0.001

Vector DNA was analyzed by PCR. ADA and purine nucleoside phosphorylase (PNP) activities were determined as described (15). ADA activities were normalized to PNP activities to correct for variability in tissue protein extraction. ADA/PNP in normal rat skin is 0.17 ± 0.04, and in cultured rat fibroblasts the ratio is 0.17 ± 0.03. ADA/PNP in uninfected control transplants was indistinguishable from those of normal skin or cultured cells. Percentage of G418^r colonies = (colonies that grew in G418)/(colonies that grew in the absence of G418) × 100. Plating efficiency of G418^r cells in G418 is 20–70% lower than that without G418; thus, even a cell population grown in G418 does not give 100% G418^r colonies (e.g., day 0). Values are the means of two or three separate explants for each period.

growth of the vector-infected cells in culture did not restore vector expression and suggest that suppression was caused by stable inactivation of vector sequences. Both the MoMLV promoter and the SV40 early promoter were subject to suppression *in vivo*.

Vector Expression in Continually Cultured Cells. Expression of the LASN vector was monitored in cultured cells that were not transplanted. LASN-infected, G418-selected rat fibroblasts were cultured in a replicating state by sequential subculture without G418. In parallel, a mixture of infected and uninfected cells was passaged as a confluent monolayer to mimic the quiescent state of transplanted cells *in vivo*. The percentage of G418-resistant cells was monitored periodically as a measure of vector expression. In both replicating and quiescent cultures, the proportion of G418-resistant cells varied no more than 3-fold over 2 mo (data not shown). This contrasts with the >1000-fold decrease in G418 resistance seen after 1 mo in animals (Table 2). Thus, suppression of expression from the transduced genes was only seen in transplanted cells.

Vector RNA in Transplanted Cells. To determine whether down-regulation of hADA and *neo* were mediated at a transcriptional or translational level, RNA was isolated from cells before transplantation and from cultured cells recovered from grafts that had been left on animals for 1 mo (Fig. 6). Although the cultures established from grafts contained between 4% and 11% vector-positive cells (Table 3), neither LTR- nor SV40-promoted transcripts were detectable. Although decreased message stability could explain this result (i.e., cells recovered from grafted tissue are altered in a way that renders vector messages differentially unstable), it is more likely that message accumulation is decreased by transcriptional mechanisms.

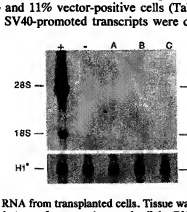


FIG. 6. RNA from transplanted cells. Tissue was removed from three animals 1 mo after transplant, and cellular RNA was prepared from cells cultured from grafts. Ten micrograms of RNA was run in each lane of an agarose gel and then transferred to nylon filters. Filters were hybridized with ^{32}P-labeled *neo* probe and then rehybridized with histone H1° probe to show equal RNA loading in each lane. Full-length LTR messages migrate just below the 28S ribosomal band, whereas the shorter SV40 message migrates at ≈18S. Lanes: A, B, and C, RNA from cells grown from grafted tissue; +, RNA from hADA-expressing nontransplanted fibroblasts; −, RNA from uninfected fibroblasts.

Reactivation of Vector Sequences. DNA methylation can suppress provirus expression, an effect that can be partially reversed by growth of cells in 5-azacytidine (18, 19). In some instances, proviruses can also be reactivated to a limited extent by using 5-bromodeoxyuridine (20). To determine whether similar mechanisms were responsible for vector inactivation in fibroblasts, cells cultured from dermal equivalent grafts were treated with 5-azacytidine (3 μM, 48 hr) or with 5-bromodeoxyuridine (20 μg/ml, 48 hr). Activation was monitored by determining the number of cells able to form colonies in G418 (Table 3). One of the three cultures (B) shows a slight increase in G418-resistant colonies after treatment with 5-azacytidine similar to the levels of reactivation seen in previous work (18, 20)—i.e., 5–60 reactivated cells per 10^6 treated cells. Although 5-azacytidine had no measurable effect in the other two cultures, the measurements are too close to the limits of detection to show a definitive lack of response. Treatment of cells with 5-bromodeoxyuridine had no detectable effect in any cultures isolated from grafts.

Two of the cultures (A and B) exhibited small increases in G418-resistant colony formation after treatment with sodium butyrate (5 mM, 48 hr). Sodium butyrate induces hyperacetylation of histones, and the resulting increase in expression from some genes is thought to occur through modification of chromatin structure (21). This result, along with the small amount of reactivation seen with 5-azacytidine, suggests that *de novo* methylation and chromatin restructuring may play small roles in vector inactivation, but the inability of these drugs to reactivate the vector sequences in most cells suggests that the major cause of inactivation remains unexplained.

DISCUSSION

These studies show profound suppression of transduced gene expression in transplanted fibroblasts. One month after transplantation, vector-encoded hADA expression was down >1500-fold. The effect was not seen during long-term culture of the cells *in vitro*, and the suppression was not reversed by recultivation of the transplanted cells. In contrast, the endogenous rat ADA gene was active continuously in the transplanted cells (Fig. 3). The vector-infected cells used were polyclonal (>10^5 infected clones) populations, and nine separately infected populations of cells were used, suggesting that suppression was independent of the virus integration site. A lack of measurable immune responses to transplanted

Table 3. G418 resistance of cells isolated from grafts after treatment with 5-azacytidine, bromodeoxyuridine, or sodium butyrate

Culture	Vector-positive cells, %	G418^r colonies per 10^6 total colonies			
		Untreated	5-AzaC	BrdUrd	Butyrate
A	11	0.7	<4.0	<4.2	3.9
B	7	0.6	36	<4.7	17
C	4	<1.1	<17	<36	<3.7
−	<0.01	<0.7	<13	<19	<4.7
+	100	3.2×10^5	2.2×10^5	4.9×10^5	5.7×10^5

Cells cultured from grafts that had remained on animals for 1 mo (A, B, C) were treated with the indicated chemicals and assayed for colony formation with and without G418. −, Uninfected fibroblasts; +, infected G418-resistant fibroblasts; numbers preceded by < indicate the sensitivity of the assay when no G418-resistant colonies were scored. BrdUrd, bromodeoxyuridine; 5-AzaC, 5-azacytidine.

cells and the high proportion of donor cells in the graft at all times suggests that selection for rare nonexpressing cells was unlikely. This was also supported by sex-mismatched transplants, where the proportion of female cells isolated from male hosts was the same regardless of whether or not the donor cells expressed hADA at the time of transplant.

Although there were no overt signs of immune response to the grafted tissue, histological analysis showed minor infiltration consistent with normal wound healing. Possibly, the process of wound healing produces inhibitory cytokines that contribute to decreased expression. Regardless of the actual mechanism involved, the absence of selection against vector-expressing cells suggests that suppression is a result of gene regulation and not of cell death.

The LASN vector has also been used to infect mouse hematopoietic stem cells (22, 23). Although evidence exists for inactivation in lymphoid cells, hADA was clearly present in blood of most transplanted mice for >5 mo, in some mice at levels equivalent to those of endogenous mouse ADA. This result demonstrates that the LASN vector can achieve long-term *in vivo* expression in some instances and that suppression occurs via a mechanism specific for certain cell types, including fibroblasts.

Perhaps regulatory elements within the LASN vector can be modified to provide expression in fibroblasts. This approach has been successfully used in murine embryonal carcinoma cells and embryo stem cells where MoMLV expression is strongly suppressed (18–20). MoMLV regulatory sequences are not functional in these cell types (24), and mutations within the MoMLV enhancer and adjacent sequences have been used to map the sequences that limit expression (25–28). With this information, several new vectors have been derived from MoMLV that are expressed in embryonal cells. Because vectors that exhibit suppression of expression *in vivo* are all based on MoMLV (5, 6), similar manipulation of the MoMLV enhancer may yield vectors capable of long-term expression in transplanted fibroblasts. These experiments point to significant problems that must be solved before gene therapy with fibroblasts can be attempted and provide an interesting model for studying regulation of gene expression *in vivo*.

Sequence information for Neo1 and Neo5 primers was kindly provided by Richard Morgan. We thank Eileen Bryant and the Fred Hutchinson Cancer Research Center Department of Cytogenetics for help with cytogenetic analysis. This work was supported by Grants AG00057, HL41212, AI19565, and DK38531 awarded by the National Institutes of Health.

1. Miller, A. D. (1990) *Blood* **76**, 271–278.
2. Garver, R. I., Jr., Chytil, A., Courtney, M. & Crystal, R. G. (1987) *Science* **237**, 762–764.
3. Selden, R. F., Skoskiewicz, M. J., Russell, P. S. & Goodman, H. M. (1987) *N. Engl. J. Med.* **317**, 1067–1076.
4. Rosenberg, M. B., Friedmann, T., Robertson, R. C., Tuszynski, M., Wolff, J. A., Breakefield, X. O. & Gage, F. H. (1988) *Science* **242**, 1575–1578.
5. St. Louis, D. & Verma, I. M. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3150–3154.
6. Palmer, T. D., Thompson, A. R. & Miller, A. D. (1989) *Blood* **73**, 438–445.
7. Wolff, J. A., Fischer, L. J., Xu, L., Jinnah, H. A., Langlais, P. J., Iuvone, P. M., O'Malley, K. L., Rosenberg, M. B., Shimohama, S. & Friedmann, T. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9011–9014.
8. Ham, R. G. (1980) *Methods Cell Biol.* **21a**, 255–276.
9. Hock, R. A., Miller, A. D. & Osborne, W. R. A. (1989) *Blood* **74**, 876–881.
10. Miller, A. D. & Buttimore, C. (1986) *Mol. Cell. Biol.* **6**, 2895–2902.
11. Miller, A. D., Trauber, D. R. & Buttimore, C. (1986) *Somatic Cell Mol. Genet.* **12**, 175–183.
12. Osborne, W. R. A. & Miller, A. D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6851–6855.
13. Palmer, T. D., Hock, R. A., Osborne, W. R. A. & Miller, A. D. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1055–1059.
14. Yannas, I. V., Burke, J. F., Orgill, D.-P. & Skrabut, E. M. (1982) *Science* **215**, 174–176.
15. Osborne, W. R. A. (1980) *J. Biol. Chem.* **255**, 7089–7092.
16. Osborne, W. R. A. & Spencer, N. (1973) *Biochem. J.* **133**, 117–123.
17. Zieverink, W. D. & Moloney, W. C. (1965) *Proc. Soc. Exp. Biol. Med.* **119**, 370–376.
18. Stewart, C. L., Stuhlmann, H., Jahner, D. & Jaenisch, R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4098–4102.
19. Jahner, D., Sthulmann, H., Stewart, C. L., Harbers, K., Lohler, J., Simon, I. & Jaenisch, R. (1982) *Nature (London)* **298**, 623–628.
20. Niwa, O., Yokota, Y., Ishida, H. & Sugahara, T. (1983) *Cell* **32**, 1105–1113.
21. Candido, E. P. M., Reeves, R. & Davie, J. R. (1978) *Cell* **14**, 105–113.
22. Kaleko, M., Garcia, J. V., Osborne, W. R. A. & Miller, A. D. (1990) *Blood* **75**, 1733–1741.
23. Osborne, W. R. A., Hock, R. A., Kaleko, M. & Miller, A. D. (1990) *Hum. Gene Ther.* **1**, 31–41.
24. Linney, E., Davis, B., Overhauser, J., Chao, E. & Fan, H. (1984) *Nature (London)* **308**, 470–472.
25. Franz, T., Hilberg, F., Seliger, B., Stocking, C. & Ostertag, W. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3292–3296.
26. Barklis, E., Mulligan, R. C. & Jaenisch, R. (1986) *Cell* **47**, 391–399.
27. Hilberg, F., Stocking, C., Ostertag, W. & Gretz, M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5232–5236.
28. Weiher, H., Barklis, E., Ostertag, W. & Jaenisch, R. (1987) *J. Virol.* **61**, 2742–2746.

TAB 13

The EMBO Journal Vol. 19 No. 21 pp. 5884–5894, 2000

Retrovirus vector silencing is *de novo* methylase independent and marked by a repressive histone code

Dylan Pannell[1,2], Cameron S.Osborne[1,3], Shuyuan Yao[1,2], Tanya Sukonnik[1], Peter Pasceri[1], Angelo Karaiskakis[1], Masaki Okano[4], En Li[4], Howard D.Lipshitz[1,2] and James Ellis[1,2,5]

[1]Programs in Developmental Biology, and Cancer and Blood Research, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, M5G 1X8, [2]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada and [4]Cardiovascular Research Center, Massachusetts General Hospital, Department of Medicine, Harvard Medical School, Charlestown, MA 02129, USA

[3]Present address: Laboratory of Chromatin and Gene Expression, Department of Developmental Genetics, Babraham Institute, Cambridge, UK

[5]Corresponding author
e-mail: jellis@sickkids.on.ca

Retrovirus vectors are *de novo* methylated and transcriptionally silent in mammalian stem cells. Here, we identify epigenetic modifications that mark retrovirus-silenced transgenes. We show that murine stem cell virus (MSCV) and human immunodeficiency virus type 1 (HIV-1) vectors dominantly silence a linked locus control region (LCR) β-globin reporter gene in transgenic mice. MSCV silencing blocks LCR hypersensitive site formation, and silent transgene chromatin is marked differentially by a histone code composed of abundant linker histone H1, deacetylated H3 and acetylated H4. Retrovirus-transduced embryonic stem (ES) cells are silenced predominantly 3 days post-infection, with a small subset expressing enhanced green fluorescent protein to low levels, and silencing is not relieved in *de novo* methylase-null [dnmt3a–/–;dnmt3b–/–] ES cells. MSCV and HIV-1 sequences also repress reporter transgene expression in *Drosophila*, demonstrating establishment of silencing in the absence of *de novo* and maintenance methylases. These findings provide mechanistic insight into a conserved gene silencing mechanism that is *de novo* methylase independent and that epigenetically marks retrovirus chromatin with a repressive histone code.

Keywords: chromatin/gene silencing/methylase/stem cells

Introduction

Transcription from retrovirus long terminal repeats (LTRs) is silenced in hematopoietic and embryonic stem (ES) cells of mammals (Jahner *et al.*, 1982; Challita and Kohn, 1994; Yoder *et al.*, 1997). This stem cell-specific gene silencing event is accompanied by heavy *de novo* cytosine methylation at CpG dinucleotides in the retrovirus LTR (Jahner *et al.*, 1982; Stewart *et al.*, 1982; Bednarik *et al.*, 1990; Challita and Kohn, 1994). It is widely, but not universally, accepted that methylation is the first step in a gene silencing mechanism that ultimately controls chromatin structure (Eden *et al.*, 1998; Razin, 1998). The developmental timing of the gene silencing event can be accounted for by the presence of a postulated stem cell-specific *de novo* methyltransferase (Lei *et al.*, 1996; Tucker *et al.*, 1996) such as the recently described murine dnmt3 proteins (Okano *et al.*, 1998). Indeed [dnmt3a–/–;dnmt3b–/–] ES cells fail to methylate newly integrated exogenous retrovirus, although partial methylation of endogenous retrovirus sequences is still observed (Okano *et al.*, 1999).

Evidence that supports a functional role for methylases in retrovirus silencing includes the demonstration that the methylation inhibitor 5-azacytidine (5-azaC) can reactivate expression of silenced murine retroviruses in embryonic carcinoma (EC) cells (Stewart *et al.*, 1982). In addition, dnmt1 null mice fail to silence the intracisternal A particle (IAP) class of endogenous retrovirus that has no extracellular infectious stage (Walsh and Bestor, 1999). These findings support a genome defense model in which *de novo* methylases have evolved in higher eukaryotes to silence foreign parasitic DNA or to direct allele-specific gene expression (Walsh and Bestor, 1999).

The action of methylases has direct and indirect effects on gene expression (Razin, 1998). Methyl groups on modified DNA directly affect the ability of some transcription factors to bind DNA, and methylases indirectly accomplish the same result by provoking condensation of chromatin structure. Such chromatin structure alterations are mediated by the binding of MeCP2 factor to methylated DNA. MeCP2 in turn interacts with a complex of several co-repressor molecules including mSin3A and the histone deacetylases HDAC1 and HDAC2 (Jones *et al.*, 1998; Nan *et al.*, 1998). Therefore, *de novo* methylases are responsible for the primary tag that marks a DNA sequence for MeCP2 binding, and recruitment of the mSin3A–HDAC complex results in histone deacetylation, which condenses chromatin and prevents transcription factor access.

Some retrovirus silencing results are not easily explained by *de novo* methylase-dependent silencing pathways. Specifically, the infectious C-type retrovirus Moloney murine leukemia virus (MoMLV) is silenced in murine EC cells 2 days post-infection, but methylation at the LTRs is not detectable until 8–16 days post-infection (Gautsch and Wilson, 1983; Niwa *et al.*, 1983; Kempler *et al.*, 1993). Unlike EC cells, more differentiated cell types such as murine erythroleukemia (MEL) cells express MoMLV-based vectors initially, but these vectors are silenced progressively by an HDAC-mediated mechanism that precedes methylation density-dependent silencing

©European Molecular Biology Organization

(Lorincz *et al.*, 2000). In addition, methylation does not correlate with silencing of endogenous genes and is not targeted to retrovirus LTRs in the sea squirt (Simmen *et al.*, 1999).

An example of a methylase-independent pathway has been described that acts to silence reporter genes linked to adeno-associated virus (AAV) sequences in HeLa or MEL cells. This pathway permits reactivation of the silenced gene by treatment with the histone deacetylase inhibitors trichostatin A (TSA) and sodium butyrate, but not by the methylation inhibitor 5-azaC (Chen *et al.*, 1997). Therefore, it appears that, in some circumstances methylase action is not a prerequisite for virus silencing. Chromatin immunoprecipitation analyses of the AAV-silenced reporter gene demonstrated that histone H4 is deacetylated at K8 and the chromatin condensed (Chen and Townes, 2000). These results suggest that H4 deacetylation is a component of the combinatorial histone code that marks silent AAV sequences. The histone code is composed of multiple post-translational modifications on histone tails that protrude from nucleosomes and on the linker histone H1 (Strahl and Allis, 2000). Modifications to H3 and H4 include acetylation of an array of lysine residues, phosphorylation of serines, and methylation. Silent chromatin is usually considered to be bound by H1 and hypoacetylated H3 and H4. The broad histone code that marks silent retrovirus vectors in embryonic or hematopoietic stem cells has not been described.

In contrast to the complex situation in mammals, methylation is not detectable in *Drosophila* (Urieli-Shoval *et al.*, 1982), and embryonic gene silencing in flies is dependent on altered chromatin structures or silencing factors. For example, the *Polycomb* group (*Pc*-G) genes are silencing factors that maintain patterns of repression established in the early embryo (Pirrotta, 1998). Reversal of such *Pc-G* silencing is accompanied by H4 hyperacetylation (Cavalli and Paro, 1999). Given the high degree of conservation of *Pc*-G and HDAC genes between eukaryotes (Pazin and Kadonaga, 1997), it is likely that gene silencing mechanisms themselves are also widely conserved. In fact, several homologs of methyl-binding proteins are conserved in *Drosophila* that do not bind methylcytosine but retain the ability to interact with HDACs, such as those in the Mi-2–NuRD complex (Tweedie *et al.*, 1999). Therefore, *Drosophila melanogaster* presents a unique system for the study of methylase-independent silencing mechanisms that are highly conserved.

MoMLV contains several *cis*-acting sequences that function as silencers. The finding that MoMLV-based vectors containing all of these elements are silenced in hematopoietic stem cells has limited their potential effectiveness for gene therapy purposes (Challita and Kohn, 1994). In an effort to relieve silencing effects, the murine stem cell virus (MSCV) (Hawley *et al.*, 1994) and HSC1 vectors (Osborne *et al.*, 1999) were developed containing mutations in the LTR direct repeat and primer-binding site, or all known silencer elements, respectively. Our assay for gene silencing takes advantage of a powerful human β-globin reporter gene regulated by the locus control region (LCR) in transgenic mice (Ellis *et al.*, 1997). The human β-globin LCR is composed of at least four erythroid-specific DNase I-hypersensitive sites (HS) (Grosveld *et al.*, 1987). These HS are nucleosome-free regions of open chromatin that are highly accessible to *trans*-acting factors (Tuan and London, 1984; Forrester *et al.*, 1986). LCR activity confers copy number-dependent expression in transgenic mice (Grosveld *et al.*, 1987; Ryan *et al.*, 1989), and transgene expression is accompanied by the establishment of HS at the LCR and the β-globin promoter (Ellis *et al.*, 1996).

Here, we show that addition of MoMLV and human immunodeficiency virus type 1 (HIV-1) sequences completely silences the LCRβ-globin reporter transgene at all integration sites. MSCV sequences also silence the LCRβ-globin reporter transgene and prevent the formation of HS at the LCR. The reduced chromatin accessibility of the silent transgene is code marked by linker H1, deacetylated H3 and acetylated H4. Retrovirus vectors also silence reporter genes in [dnmt3a–/–;dnmt3b–/–] ES cells and transgenic *Drosophila*, showing that retrovirus silencing does not require *de novo* methylases. These independent and complementary data demonstrate that gene silencing of retrovirus sequences occurs in mice and *Drosophila*, and are consistent with the existence of a conserved retrovirus gene silencing mechanism that involves repressive histone code marks.

Results

Retrovirus vectors silence LCRβ-globin transgenes in mice

To determine whether retrovirus sequences can silence a powerful LCRβ-globin transgene in mice, we linked MoMLV and MSCV sequences to an antisense orientation of the BGT14 reporter cassette, which expresses at 16–71% per transgene copy and at all integration sites (Ellis *et al.*, 1997). This orientation is preferred in β-globin retrovirus vectors in order to preserve the introns during retrovirus replication (Lebouich *et al.*, 1994). The BGT14 reporter gene contains a 3.0 kb LCR composed of 5′HS2–4 that is linked to a 4.2 kb human β-globin gene regulated by its 815 bp promoter. Its ability to express β-globin reproducibly and consistently in transgenic mice makes it an ideal reporter for silencing activity. The microinjection fragments used in the transgenic mouse silencing assays are described in Figure 1. BGT25 contains 1.5 kb of MoMLV 5′ LTR and extended packaging sequences, while BGT32 contains the equivalent 1.5 kb sequence derived from MSCV (Figure 1A).

The BGT25 and BGT32 transgenes were introduced into mice, and founder fetuses were collected at embryonic day 15.5. Transgenic animals were identified by Southern blot analysis with the βivs2 probe on fetal head DNA. Transgene copy number and intactness were determined by DNA digestion with restriction enzymes that reveal end fragments (*Eco*RI and *Bam*HI for copy number) or internal fragments (for intactness). S1 nuclease analysis of fetal liver RNA showed complete silencing of the human β-globin transgene in all BGT25 (6/6) transgenic mice, and complete silencing in seven of eight BGT32 transgenic animals (Figure 1B). High level expression of a human β-globin transgene is detectable in the positive control fetal liver RNA from the μD14 transgenic mouse line (Ellis *et al.*, 1996). Therefore, silencing mediated by MoMLV (McCune and Townes, 1994) and MSCV

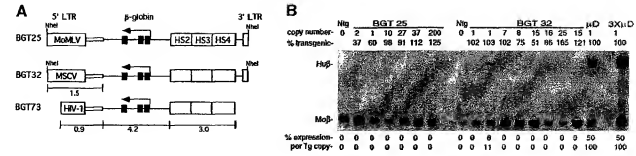


Fig. 1. Silencing by C-type retrovirus sequences in transgenic mice. (A) Structure of LCRβ-globin reporter transgenes. MSCV contains several mutations in viral silencers. HS2, HS3, HS4: human β-globin LCR elements. (B) S1 nuclease analysis of globin expression in RNA of 15.5 day fetal livers showing that BGT25 (MoMLV) and BGT32 (MSCV) transgenes are completely silenced, with one exception. Huβ, human β-globin protected probe fragment; Moβ, mouse β-major protected probe fragment; Ntg, non-transgenic; μD, one-copy μD14 microlocus transgenic line; 3XμD, probe excess control.

sequences is dominant to the activating functions of the β-globin LCR.

Silenced transgenes are in inaccessible chromatin

To examine the chromatin structure of the LCR and β-globin promoter in a silenced BGT32 transgene, we generated a silenced mouse line and performed HS mapping (Forrester *et al.*, 1990) on F_1 fetal liver nuclei (Figure 2). Digestion with *Eco*RI and hybridization with the βivs2 probe did not detect any HS in the transgene β-globin promoter or LCR. In contrast, HS in the endogenous mouse β-globin LCR (Ellis *et al.*, 1996) in the same samples was detected. Hence, the chromatin of the silenced transgene is inaccessible to the factors that cause HS formation, and this inaccessible structure could itself prevent activation of the reporter transgene.

Silencing does not occur by transcriptional interference

Transgene silencing by retrovirus sequences could be due to transcriptional interference initiated from the viral 5′ LTR promoter (Cullen *et al.*, 1984). Actively transcribed 5′ LTRs possess an HS at both the initiation site and the enhancer, but the 3′ LTR forms only the enhancer HS (Thompson and Fan, 1985; Rasmussen and Gilboa, 1987). We used the same DNase I-treated samples described above to search for HS formation on the viral LTR of the MSCV-silenced transgene. Digestion with *Bam*HI and hybridization with the βivs2 probe demonstrated no HS elements on the β-globin 3′ enhancer, nor on the 5′ LTR (Figure 2). These data are inconsistent with transgene silencing mediated by transcriptional interference. Rather than being in an expressed conformation, the retrovirus sequences are themselves in an inaccessible chromatin structure.

Silenced LCRβ-globin transgenes have a repressive histone code

Inaccessible chromatin structures present on silent genes are believed to have a combinatorial histone code that includes bound H1 and hypoacetylated H3 and H4. Phosphorylated Ser10 on H3 is also present on condensed mitotic chromosomes (Wei *et al.*, 1999), suggesting that it may also be found on silent chromatin. To discover the broad histone code of MSCV-silenced retrovirus transgenes, we performed chromatin immunoprecipitation (ChIP) assays on day 15.5 fetal liver cells explanted from the BGT32 transgenic mouse line. Cross-linked chromatin was immunoprecipitated with commercially available antibodies that specifically recognize biacetylated H3 (H3-Ac), phospho-Ser10 H3 (H3-P), hyperacetylated H4 (H4-Ac) or unmodified H1. Specific transgene DNA fragments in the input, pellet and supernatant DNA fractions were detected by Southern hybridization after equal loading onto slot-blots. A representative ChIP experiment for each antibody and probe is shown in Figure 3. Probes for both the human 5′HS3 LCR element and the silent human β-globin promoter did not detect signals in the H3-Ac pellet, but visible signals of increasing intensity were observed in the H4-Ac and H1 pellet fractions. A probe for the expressed endogenous mouse β-major promoter detected moderate H3-Ac and

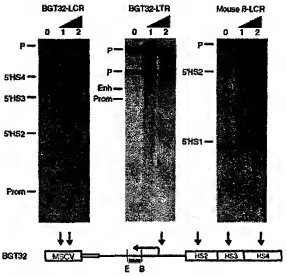


Fig. 2. DNase I-hypersensitive site mapping of 15.5 day fetal liver nuclei from a silenced four-copy BGT32 line showing that the transgene β-globin promoter, LCR and retrovirus LTR elements are in an inaccessible chromatin conformation. The endogenous mouse β-globin LCR elements in the same samples are hypersensitive. Expected transgene HS are shown below the vertical arrows. E, *Eco*RI; B, *Bam*HI; the black rectangle indicates the βivs2 probe.

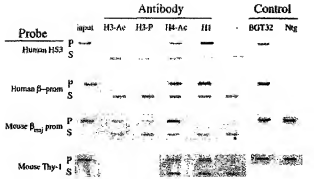


Fig. 3. ChIP assays detect a repressed histone code on the MSCV-silenced LCRβ-globin reporter in BGT32 transgenic mice. ChIP was performed with antibodies that detect H3-Ac, H3-P, H4-Ac and H1. Histone modifications at specific DNA fragments were detected by slot-blot hybridization with the probes listed on the left.

strong H4-Ac signals, but little or no H1 signal. These data suggest that the histone code of the silent BGT32 transgene includes abundant H1, deacetylated H3 and acetylated H4. In contrast, the histone code for the expressing mouse β-major promoter includes depleted H1, acetylated H3 and hyperacetylated H4. This expressing histone code also differs from the acetylated H4 and bound H1 found on the silent mouse *thy-1* gene. ChIPs with the H3-P antibody showed more variability, with moderate signals detected with the human 5′HS3 probe and mouse β-major promoter probe, but not with the human β-globin promoter probe. These results suggest that H3-P does not in general correlate with silent chromatin states.

To identify the specific histone code marks that correlate best with the silenced transgene state, the ChIP results were quantified using a PhosphorImager. The intensity of the band in the no antibody (–) slot was subtracted from that of each pelleted band, and this normalized intensity was expressed as a percentage of the normalized input DNA. To calculate relative fold differences in histone code marks, the mean average percentage of input values obtained from at least two independent immunoprecipitations was calculated, and fold differences were determined with respect to the expressed mouse β-major gene (Table I). These calculations show that the silent human β-globin promoter is marked by a 9.1-fold enrichment of H1, an 8.8-fold depletion of H3-Ac, but only a 2.3-fold depletion of H4-Ac. Similar code marks tag the human 5′HS3 element, although depletion of H4-Ac is increased to 6.1-fold. We conclude that the most dramatically altered histone code marks at both 5′HS3 and the silent transgene promoter are abundant H1 and deacetylated H3. These data are the first description of the histone code of retrovirus-silenced transgenes in mice, and are consistent with the formation of inaccessible chromatin structures. The histone code described above suggests that retrovirus silencing in transgenic mice involves either the action of an HDAC that has a greater impact on H3 acetylation ratios than it does on H4, and/or reflects reduced accessibility by H3-specific HATs to the transgene chromatin.

Table I. Quantitation of ChIP data on individual histone code marks in silenced BGT32 transgenic mice compared with the expressed mouse β-major globin gene

	H3-Ac	H3-P	H4-Ac	H1
Hβ-HS3	–9.5× (5.0)	–2.5× (11.2)	–6.1× (69.5)	7.4× (136.1)
Hβ-Prom	–8.8× (5.4)	–284× (0.1)	–2.3× (181.3)	9.1× (167.7)
Mβmaj-Prom	1.0× (47.4)	1.0× (28.4)	1.0× (422.7)	1.0× (18.5)
MThy1	ND	ND	–7.0× (60.3)	4.7× (86.1)

Figures in parentheses denote the mean average percentage of input DNA values for each pellet fraction from at least two independent ChIP experiments.
ND, not done.

Retrovirus vectors silence GFP reporters in de novo methylase-null ES cells

Could *de novo* methylases play a crucial role in targeted recognition of the retrovirus sequences and subsequent recruitment of the postulated HDAC as suggested in the genome defense model? Previous gene targeting procedures have generated ES cells that are compound homozygous null mutants for the *de novo* methylases dnmt3a and dnmt3b. In particular, the 7aabb [dnmt3a–/–; dnmt3b–/–] ES cell line has residual maintenance methylase activity, but is incapable of *de novo* methylation of newly infected exogenous MoMLV containing a SupF marker (Okano *et al.*, 1999). To determine the importance of *de novo* methylases in retrovirus silencing, we compared expression of several retrovirus vectors containing an internal phosphoglycerate kinase (PGK)–enhanced green fluorescent protein (EGFP) reporter cassette in 7aabb and wild-type J1 ES cells over time. NIH 3T3, J1 and 7aabb cells were infected simultaneously in duplicate with each of the virus stocks, and GFP expression monitored over time by flow cytometry. Infections with MSCV–EGFP virus generated 30% GFP$^+$ NIH 3T3 cells versus ~7% GFP$^+$ J1 and 7aabb cells at all time points (Figure 4). Mean fluorescence in expressing NIH 3T3 cells at day 10 was 1092 compared with 147–180 in the expressing ES cells. Therefore, the apparent titer in ES cells was 4.2-fold less than in NIH 3T3 cells, and mean fluorescence decreased by 7.4-fold. Assuming similar infection frequencies in NIH 3T3 and ES cells, these data suggest that retrovirus silencing is established by day 3 with equal efficiency in J1 and 7aabb ES cells. Infections with the LNCX–EGFP virus produced almost identical data (Figure 4, earlier time points not shown), indicating that a subset of infected ES cells permit low level expression of both wild-type and MSCV vectors.

We then tested whether the SIN deletion in the HSC1–EGFP vector relieves silencing in ES cells. HSC1–EGFP infection generated 24% GFP$^+$ NIH 3T3 cells compared with 8% GFP$^+$ J1 and 7aabb cells at all time points, with mean fluorescence at day 10 of 517 in NIH 3T3 compared with 342–376 in the ES cells (Figure 4, earlier time points not shown). The apparent titer in ES cells was 3-fold less than in NIH 3T3 cells, and mean fluorescence decreased by only 1.5-fold, remaining higher than either MSCV- or LNCX virus-transduced ES cells. As the percentage of expressing cells and their mean fluorescence levels were indistinguishable in J1 and 7aabb cells for LNCX, MSCV and HSC1 retrovirus vectors, we

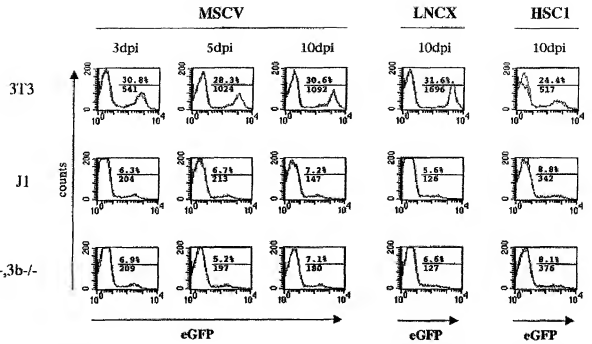


Fig. 4. Flow cytometry demonstrates that retrovirus vectors are silenced in wild-type and *de novo* methylase-null ES cells. The PGK–EGFP cassette was delivered using MSCV, LNCX and HSC1 retrovirus vectors to NIH 3T3, wild-type ES cells (J1) and [dnmt 3a–/–;dnmt3b–/–] ES cells, and flow cytometry was performed 3, 5 and 10 days post-infection. Each infection was performed in duplicate (blue and red lines) and the data merged. The average percentage of GFP+ cells is shown above the line, and mean fluorescence below the line, indicating the gate used.

conclude that retrovirus silencing is established with equal efficiency in the presence or absence of the *de novo* methylases dnmt3a and dnmt3b.

Drosophila reporter transgenes detect conserved silencing mechanisms

If chromatin structure alterations are a primary cause of retrovirus silencing, then the factors involved may be conserved in other organisms that lack methylases. To this end, we employed a transgenic *Drosophila* reporter system developed for the characterization of *Polycomb*-responsive elements (PREs). The P[U/I5] reporter transgene contains GAL4-UAS sites upstream of reporter *lacZ* and mini-*white* genes (Figure 5). In the presence of GAL4, the *lacZ* reporter gene is highly expressed. In contrast, the P[UZS1] transgene contains the mouse H19 imprinting center that functions as a silencer in *Drosophila* and prevents *lacZ* reporter expression (Lyko *et al.*, 1997). Silencing in this system has been attributed to chromatin condensation effects that block GAL4 binding (Zink and Paro, 1995). Hence, this reporter system is ideal for investigating conserved silencing mechanisms.

Retrovirus vectors silence transgenes in Drosophila

To test whether C-type retrovirus sequences are silenced in the absence of methylases but the presence of conserved silencing factors, we linked an MSCV target sequence to the GAL4-UAS-*lacZ* and mini-*white* reporter genes (Figure 5) and introduced the resulting P[UZVSR] construct into *Drosophila*. Seven independent P[UZVSR] transgenic lines containing MSCV sequences were generated by P-element-mediated germline transformation. The

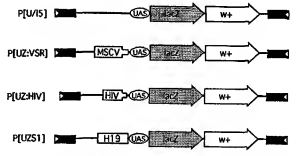


Fig. 5. Structure of reporter transgenes used in *Drosophila* to assay for conserved silencing mechanisms. The construct P[U/I5] contains GAL4-UAS sites regulating the *lacZ* gene and the mini-*white* transformation marker. In the presence of GAL4, the *lacZ* gene is expressed to high levels. P[UZVSR] contains MSCV sequences, P[UZHIV] contains HIV-1 sequences and P[UZS1] contains the mouse H19 imprinting center, which functions as a silencer in *Drosophila* and blocks *lacZ* activation by GAL4.

eye color of these lines ranged from orange to yellow, consistent with reduced to very low mini-*white* expression (Figure 6A). None of the flies demonstrated variegated eye color, and flies with white eyes were assumed to be non-transgenic. We crossed the P[UZVSR] lines to the 1032.hx line that expresses GAL4 in salivary glands and stained late stage embryos of the line VSR:B for β-galactosidase expression (Figure 6B). VSR:B expresses β-galactosidase to very low but uniform levels in the salivary glands, and the non-silenced P[U/I5-2] control expresses to high levels as expected (Figure 6B, upper panels). Dissected salivary glands from third instar larvae also show uniformly reduced β-galactosidase expression that is restricted primarily to the nucleus in lines VSR:AK and AG, and very low levels in VSR:B (Figure 6B, lower panels).

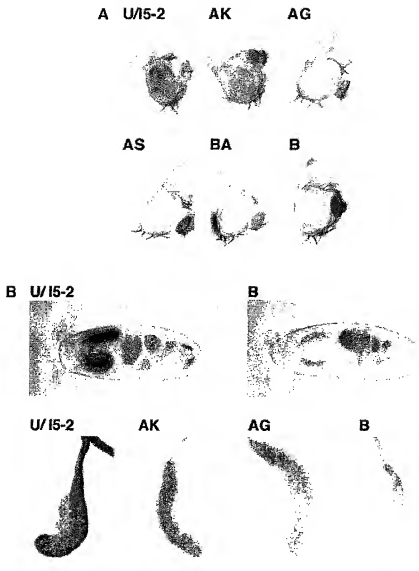


Fig. 6. MSCV sequences silence linked mini-*white* and *lacZ* reporter genes in *Drosophila*. (**A**) Heads of age-matched adult flies showing that the *white* gene is expressed at high levels in non-silenced P[U/I5] control flies, at intermediate levels in lines VSR:AK and AG, and at low levels in lines VSR:AS, BA and B. (**B**) Stage 16 embryos stained with X-Gal show high level β-galactosidase expression in salivary glands of P[U/I5] but low levels in VSR:B (upper panels). Salivary glands dissected from wandering third instar larvae stained with X-Gal show high level β-galactosidase expression in P[U/I5], intermediate levels in VSR:AK and AG, and low levels in VSR:B (lower panels).

Levels of β-galactosidase were quantified in each line from homogenates of third instar larvae. In the presence of GAL4, β-galactosidase enzymatic activity was significantly reduced in all seven lines (Figure 7A), and lines VSR:B and VSR:BA were equivalent to the control silenced P[UZS1] line. We also created four additional non-silenced P[U/I5] control lines, all of which had red eyes (data not shown) and expressed significant β-galactosidase activity in the presence of GAL4 ranging from 70 to 99% of the levels of the published U/I5-2 line (Figure 7B). These results demonstrate that the reporter transgene is minimally subject to position effects, and argue that *lacZ* and mini-*white* silencing are specific consequences of MSCV sequences.

To examine the potential involvement of transcriptional interference from the MSCV promoter, we performed RT–PCR (Figure 7C) to detect transcripts that initiate at the MSCV promoter (M1 primer pair) or upstream of the MSCV promoter (M2 primer pair). We observed no expression from the MSCV promoter with the M1 primers in line VSR:BA, and detected GAL4-independent upstream transcripts with both M2 and M1 primers in line VSR:B. As expected, RT–PCR with the L1 primer pair detected GAL4-dependent *lacZ* mRNA in both these lines. These results are inconsistent with transcriptional interference from the MSCV promoter. We therefore conclude that MSCV sequences silence both closely linked *lacZ* and more distant *white* genes via a methylase-independent pathway in transgenic *Drosophila*.

Lentivirus vectors silence transgenes in mice and Drosophila

To determine whether transcriptional silencing occurs in other C-type retroviruses that have diverged from the

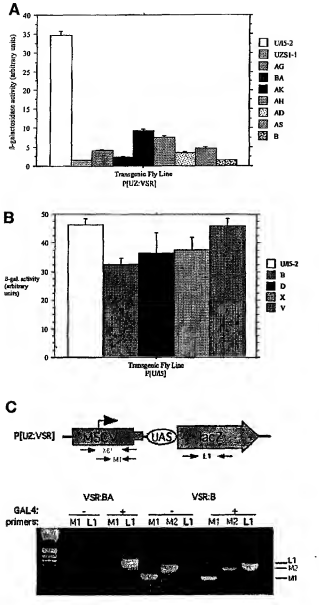


Fig. 7. MSCV sequences significantly reduce β-galactosidase enzymatic activity in transgenic *Drosophila*. (**A**) The P[UZVSR] lines express a range of 4–26% of the non-silenced P[U/I5-2] line. (**B**) The P[U/I5] lines exhibit only minor position effects and express from 70 to 99% of the levels of the control P[U/I5-2] line. Error bars indicate the standard error from the mean values shown. (**C**) RT–PCR with primer sets (above) show no MSCV transcription (below) in line VSR:BA, or transcription initiated upstream of MSCV in line VSR:B.

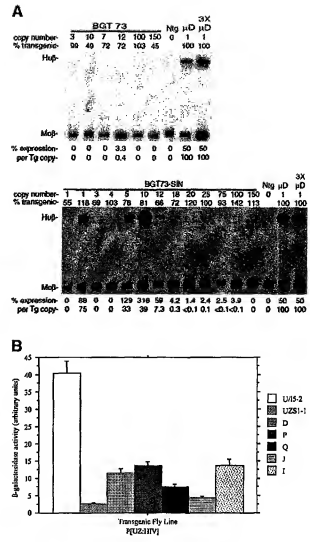


Fig. 8. HIV-1 sequences silence reporter transgene expression in mice and *Drosophila*. (**A**) S1 nuclease analysis of globin expression in RNA of 15.5 day fetal livers showing that BGT73 (HIV-1) transgenes are completely silenced in mice, while silencing is ameliorated but not eliminated in the HIV-SIN construct. (**B**) HIV-1 sequences significantly reduce β-galactosidase enzymatic activity in transgenic *Drosophila*. The P[UZHIV] lines express a range of 12–31% of the non-silenced P[U/I5-2] line.

murine retroviruses tested above, we inserted the 5′ LTR and minimal packaging site of HIV-1 as a 0.9 kb fragment into the LCRβ-globin reporter transgene to create the BGT73 construct (Figure 1A). Mice bearing the BGT73 construct were generated and characterized as described above. S1 nuclease analysis of 15.5 day fetal liver RNA (Figure 8A, upper panel) shows that the β-globin transgene is completely silenced in all the BGT73 mice obtained (6/6). We also purified an HIV-SIN injection fragment that deletes the U3 region to −18 bp (Zufferey *et al.*, 1998) directly from the BGT73 plasmid and tested this construct in transgenic mice. S1 nuclease analysis of 15.5 day fetal liver RNA (Figure 8A, lower panel) shows that the β-globin transgene is completely silenced in 7/13 HIV-SIN mice. We conclude that the HIV-SIN construct ameliorates transgene silencing but retains some novel silencer elements.

We obtained five transgenic *Drosophila* lines containing the P[UZHIV] construct (Figure 5A) to examine whether HIV-1 sequences are silenced by a methylase-independent pathway. Eye color of these lines was not reduced observably, similar to the P[UZS1] H19 control but in contrast to the P[UZVSR] lines that contain MSCV. In the presence of GAL4, β-galactosidase activity was reduced in all five P[UZHIV] lines (Figure 8B) to levels comparable with that by MSCV. These results show that although HIV-1 silencing has been correlated with high methylation levels in mammals (Bednarik *et al.*, 1990), it

can also occur by a conserved pathway in the absence of methylases.

Discussion

One of the first epigenetic models of mammalian gene silencing was proposed to account for the behavior of C-type retroviruses in murine stem cells and embryos (Jahner *et al.*, 1982). The genome defense model specifically proposes that *de novo* methylases are responsible for recognizing newly integrated viral DNA, thereby creating binding sites for methyl-binding proteins that recruit chromatin remodeling complexes such as HDACs. Some features of this model have not been tested *in vivo*. Here, we demonstrate that chromatin on silent retrovirus vectors *in vivo* is inaccessible and is marked differentially by a repressive histone code in transgenic mice. We also show that the efficiency of retrovirus silencing *in vivo* is not altered in *de novo* methylase-null ES cells, and that a methylase-independent retrovirus silencing pathway is conserved in *Drosophila*. These data argue that chromatin modifications play a role in C-type retrovirus silencing as predicted, and that *de novo* methylases are not required for retrovirus silencing by stem cells.

Role of chromatin structure in retrovirus silencing

We sought to obtain direct *in vivo* evidence that inaccessible chromatin is established by retroviruses in embryonic cells. To this end, we tested C-type retrovirus silencing in transgenic mice. An LCRβ-globin cassette was chosen as the reporter gene because it opens chromatin and expresses at all ectopic integration sites tested. MoMLV, MSCV and HIV-1 vectors dominantly silence transcription by this reporter, indicating that viral sequences establish inaccessible chromatin that spreads to prevent *trans*-acting factor access to the regulatory elements on the reporter transgene. Direct evidence of such chromatin structure was obtained by demonstrating that HS do not form on the retrovirus LTR in BGT32 transgenic mice. Moreover, this inaccessible structure spreads over 4 kb to prevent HS formation on the β-globin reporter LCR, promoter and enhancer elements.

Inaccessible chromatin structure may be formed by deacetylation of the nucleosomal core histones. Chromatin immunoprecipitation experiments on the silent transgene in BGT32 fetal liver cells confirm that retrovirus silencing is marked differentially by a histone code of deacetylated H3, acetylated H4 and abundant H1. In contrast to the silenced LCRβ-globin transgene, the endogenous mouse β-major promoter has a histone code composed of acetylated H3 and hyperacetylated H4. Our finding that an 8.8-fold depletion in H3 acetylation levels best distinguishes silent from expressing globin genes agrees with a similar report on the importance of H3 acetylation in the human β-globin locus (Schubeler *et al.*, 2000). Overall, our data provide direct evidence for the establishment and spread of inaccessible chromatin structures from silent retroviruses, and the ChIP results describe a repressive histone code that localizes to silent retroviruses in mouse embryos.

Role of de novo methylases in retrovirus silencing

The transgenic mouse results correlate chromatin structure and histone modifications with retrovirus silencing but do not address the importance of *de novo* methylases in this process. To this end, we examined silencing of an internal PGK–EGFP reporter transgene by MoMLV, MSCV and HSC1 retrovirus vectors. Infection of wild-type ES cells and *de novo* methylase-null [dnmt3a–/–;dnmt3b–/–] ES cells with these vectors failed to reveal any differences in GFP silencing assayed by flow cytometry. Silencing was established by day 3 prior to completion of *de novo* methylation in wild-type cells, and was maintained essentially unaltered until day 10 when methylation is complete in wild-type cells (Okano *et al.*, 1999). Some proviruses in ES cells do escape complete silencing and express low levels of GFP. These data suggest occasional position-dependent proviral expression, and perhaps reflect integration into highly transcribed regions of the genome. Nevertheless, GFP fluorescence from these sites is reduced 7.4-fold relative to the levels in NIH 3T3 cells. We conclude that the *de novo* methylases dnmt3a and dnmt3b are not required for retrovirus silencing in undifferentiated stem cells.

Evidence for a conserved silencing mechanism

Transgenic *Drosophila* lack both *de novo* and maintenance methylases, but share other conserved gene silencing pathways with mammalian cells. Therefore, we created transgenic *Drosophila* containing MSCV and HIV-1 vector sequences to examine their ability to silence a Gal4-UAS-*lacZ* reporter gene by methylation-independent conserved pathways. In the absence of any silencer element, five P[UI/5] transgenic fly lines express β-galactosidase to high and reproducible levels with only minimal position effects. Addition of MSCV sequences to the transgene decreased β-galactosidase expression to 4–26% of the levels with uniform expression of *lacZ* in salivary glands and also reduced expression of the mini-*white* transformation marker in eyes. Silencing is not mediated by transcriptional interference from the MSCV retrovirus LTR in mice or flies. It is notable that MSCV does not completely silence *Drosophila* transgenes. The reasons for this are unclear, but may be related to the exclusion of completely silenced transgenes by selection for fly lines that express mini-*white*. The reduced levels of *lacZ* expression in flies mimic expression levels observed in the GFP$^+$ MSCV-transduced ES cells. Although we cannot formally exclude the possibility that silencing occurs by two different mechanisms in mice and flies, it is unlikely given the evidence for *de novo* methylase independence in both organisms.

Taken together, these data implicate a conserved mechanism that silences MSCV in the absence of *de novo* methylases in ES cells and in the methylase-free environment of *Drosophila*. We infer that this putative pathway is not restricted to murine retroviruses because HIV-1 sequences also reduce β-galactosidase expression to 12–31% of the levels in transgenic *Drosophila*, completely silence the LCRβ-globin reporter transgene in mice, and a self-inactivating HIV-1 vector ameliorates transgene silencing in mice to the same degree as the HSC1 murine retrovirus. The involvement of highly conserved factors would suggest that silencing is not

directed specifically at retroviruses, but rather reflects a novel use of a developmentally important gene silencing mechanism.

As *de novo* methylases are not essential for silencing in undifferentiated ES cells and *Drosophila*, we postulate that methylation is either a consequence of silencing, or is a redundant parallel pathway that is not essential with regard to C-type retroviruses but functions to tag IAP retroviruses or other foreign CpG-rich DNA elements (Walsh and Bestor, 1999). We propose that the conserved pathway is responsible for causing C-type retrovirus silencing, and that this pathway ultimately dictates the repressive histone code described here by invoking HDACs or HATs that preferentially deplete H3 acetylation levels and by loading H1.

Whether these repressive histone code marks cause silencing or are a consequence of the conserved mechanism has not been determined. Lorincz *et al.* (2000) have shown that extinction of retrovirus expression in MEL cells can be overcome initially by TSA alone but requires 5-azaC at later time points. Our preliminary ChIP studies show that TSA treatment of explanted BGT32 transgenic fetal liver cells partially restores H3 acetylation levels but fails to displace H1, and β-globin transcription is not reactivated (S.Yao, C.S.Osborne and J.Ellis, unpublished data). In addition, GFP expression in transduced ES and F9 EC cells is not enhanced measurably by TSA and/or 5-azaC treatment (T.Sukonnik and J.Ellis, unpublished data). These results are consistent with the original 5-azaC treatments that show limited reactivation of retrovirus expression in murine EC cells to <1% of the level of non-silenced control cells (Stewart *et al.*, 1982). We therefore suggest that reactivation in embryonic cells will only occur once H1 is displaced, and that the above drugs are not capable of this function in our hands.

Implications for gene therapy

Gene therapy trials indicate that retrovirus vector silencing occurs in hematopoietic stem cells derived from human bone marrow or cord blood (Weinberg and Kohn, 1998). Although it is not clearly established that the same stem cell-specific pathway is involved, dissection of retrovirus silencing mechanisms in transgenic mice, ES cells and *Drosophila* may suggest strategies to overcome this block to sustained expression of therapeutic genes in transduced hematopoietic stem cells.

Materials and methods

Plasmid construction

BGT25 contains the 7.1 kb LCRβ-globin *NotI–EcoRV* fragment from BGT14 (Ellis *et al.*, 1997) in the 4.7 kb *EcoRI–NcoI* p141 vector (Leboulch *et al.*, 1994) that includes the MoMLV backbone. BGT32 contains the same 7.1 kb LCRβ-globin fragment in the 5.2 kb *ClaI–HpaI* MSCVneoEB (Hawley *et al.*, 1994) plasmid backbone. BGT73 contains the 0.9 kb *DraI–NruI* HIV-1 5′ LTR and packaging signal fragment from pHR′ (Naldini *et al.*, 1996) cloned into the *NdeI* site 3′ of the β-globin gene in BGT14. MSCV-PGK–EGFP plasmid DNA was provided by P.Leboulch. The PGK–EGFP cassette was also subcloned into LNCX (Clontech) and HSC1 vectors. P[UZVSR] contains the 5′ LTR and gag sequences of MSCVneoEB cloned as a 1.5 kb *NheI–HpaI* fragment in p[UZ] (Lyko *et al.*, 1997). P[UZHIV] contains the 0.9 kb HIV-1 fragment cloned in p[UZ].

Transgenic mice

Transgenic mice were produced in FVB fertilized eggs as previously described (Ellis *et al.*, 1997). BGT25, BGT32 and BGT73 were gel purified as 8.5 kb *NheI*, 8.6 kb *NheI* and 8.0 kb *XmnI–NheI* fragments, respectively. The 7.4 kb BGT73-SIN injection fragment was purified after digestion of BGT73 plasmid DNA with *PvuII* and *Eco*RV. Day 15.5 post-injection, fetal mice were dissected and DNA extracted from head tissue, while the fetal liver was saved frozen in two halves for future analyses. Transgenic fetuses were identified by Southern blot analysis of fetal head DNA hybridized with the βivs2 probe, and transgene copy number determined using a Molecular Dynamics PhosphorImager. The presence of the transgene in each fetal liver was determined by Southern blot analysis of fetal liver DNA digested with *AccI*, which cuts twice within the human β-globin gene (Ellis *et al.*, 1997).

RNA analysis

Day 15.5 fetal liver RNA was extracted using Trizol reagent (Gibco-BRL); 1 μg was hybridized to [γ-^{32}P]ATP-labeled double-stranded probes, and digested with 75 U of S1 nuclease (Boehringer Mannheim), and protected bands on a 6% sequencing gel were quantitated on a Molecular Dynamics PhosphorImager as described (Ellis *et al.*, 1996).

RT–PCR was performed on third instar *Drosophila* larvae RNA extracted with Trizol using the Superscript One Step RT–PCR system (Gibco-BRL) and the listed primers for the following products: L1 (5′-GAGCCTGCTAAAGCAAAAAGAAGTCACC-3′ and 5′-CGTAACCGTGCATCTGCCAGTTTGAGG-3′), M1 (5′-CCGGTACCCGTATTCCCAAT-3′ and 5′-CGCAAACATTAGATGCCGGC-3′) and M2 (5′TGTTCGCGCGCTTCTGCTCCCCG3′ and 5′CGCAAACATTAGATGCCGGC3′).

Chromatin immunoprecipitation

The ChIP assay was performed on day 15.5 fetal liver cell suspensions using kits (Upstate Biotechnology) as described. Chromatin was cross-linked *in vivo* with 1% formaldehyde at 37°C for 10 min, cells washed once with phosphate-buffered saline (PBS) plus EDTA-free complete proteinase inhibitor (Boehringer Mannheim), resuspended in 200 μl of SDS lysis buffer with proteinase inhibitor on ice for 10 min, and sonicated four times (10 s pulses) at 35% output. After centrifugation, the soluble chromatin in the supernatant was either frozen in liquid nitrogen or used immediately by diluting 200 μl of soluble chromatin into 5.5 ml of ChIP dilution buffer with proteinase inhibitors for pre-cleaning with 200 μl of salmon sperm DNA/protein A–agarose slurry. An input control of 0.5 ml of supernatant was set aside. The remaining 5 ml were split into five Eppendorf tubes. Antibodies (Upstate Biotechnology) including anti-acetylated H3, anti-phospho-H3, anti-acetyl-H4 and anti-H1 were added to four tubes, with one tube for the no antibody control. Immunoprecipitations were performed overnight and washed as described; 0.5 ml of the supernatant of each sample was kept for unbound DNA isolation, and the bound DNA was eluted from the beads, the cross-links reversed and the DNA extracted. A 5 μg aliquot of DNA from each fraction was transferred to a slot-blot for hybridization with random-primed radiolabeled probes recognizing the human β-globin 5′ *HS3* region (850 bp *SacI–PvuII* fragment), human β-globin promoter plus exon 1/exon 2 cDNA (520 bp *AccI* fragment), endogenous mouse β-globin promoter plus exon 1 and 2 (495 bp *NcoI* fragment) and endogenous mouse Thy-1 gene (1.1 kb *AccI* fragment). The slot-blot result was imaged by PhosphorImager, and quantitated with IQMac v1.2 software.

Retrovirus infection and flow cytometry

Virus was harvested from transfected PA317 producer cell populations, and 2 ml of undiluted virus plus polybrene used to infect 1×10^5 cells. GFP expression was detected by flow cytometry using a FacScan (Becton Dickinson) after gating to exclude dead cells.

Transgenic flies

Transgenic flies were raised at 25°C on standard medium unless specified otherwise. Transgenic flies were obtained using *white*^{1118} as host and pHSπ as an exogenous source of transposase. Transformants were screened on the basis of their ability to rescue *white* eye color.

β-galactosidase expression

Quantitation of β-galactosidase activity was performed in quadruplicate for each transgenic line using a chlorophenol red β-D-galactopyranoside (CPRG) assay as described (Lyko *et al.*, 1997). One male and one female wandering third instar larvae were homogenized, aliquoted into a cuvette containing 1 mM CPRG, and ΔOD_{574}/min/mg whole protein determined

over the course of an hour. β-galactosidase *in situ* staining was performed on embryos that were fixed, washed and stained in X-Gal for 5 h at 37°C in the dark. Salivary glands from wandering third instar larvae were dissected and stained for 3 min with X-Gal as described (Zink and Paro, 1995). All lines shown were processed and photographed simultaneously.

Acknowledgements

We thank R.Paro and F.Lyko for providing advice, flies and plasmids; D.Allis for ChIP advice; and P.Leboulch, R.Hawley and A.Bukovsky for retrovirus and lentivirus vectors. T.Davies and A.Bashirullah provided advice on β-galactosidase staining. S.Soodeen-Karamath, G.Knowles and L.Wei provided assistance with ES cell culture, flow cytometry and transgenic mice, respectively. This work was supported by grants from the Medical Research Council (MRC) of Canada to J.E. and to H.D.L., the Bayer/CRCS Research and Development Fund to J.E., an Ontario Thalassemia Foundation Fellowship to C.S.O., and Hospital for Sick Children Foundation, OGS and MRC Doctoral Research Awards to D.P.

References

Bednarik,D.P., Cook,J.A. and Pitha,P.M. (1990) Inactivation of the HIV LTR by DNA CpG methylation: evidence for a role in latency. *EMBO J.*, **9**, 1157–1164.

Cavalli,G. and Paro,R. (1999) Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science*, **286**, 955–958.

Challita,P.M. and Kohn,D.B. (1994) Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*. *Proc. Natl Acad. Sci. USA*, **91**, 2567–2571.

Chen,W.Y. and Townes,T.M. (2000) Molecular mechanism for silencing virally transduced genes involves histone deacetylation and chromatin condensation. *Proc. Natl Acad. Sci. USA*, **97**, 377–382.

Chen,W.Y., Bailey,E.C., McCune,S.L., Dong,J.Y. and Townes,T.M. (1997) Reactivation of silenced, virally transduced genes by inhibitors of histone deacetylase. *Proc. Natl Acad. Sci. USA*, **94**, 5798–5803.

Cullen,B.R., Lomedico,P.T. and Ju,G. (1984) Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis. *Nature*, **307**, 241–245.

Eden,S., Hashimahony,T., Keshet,I. and Cedar,H. (1998) DNA methylation models histone deacetylation. *Nature*, **394**, 842.

Ellis,J., Tan-Un,K.C., Harper,A., Michalovich,D., Yannoutsos,N., Philipsen,S. and Grosveld,F. (1996) A dominant chromatin-opening activity in 5′ hypersensitive site 3 of the human β-globin locus control region. *EMBO J.*, **15**, 562–568.

Ellis,J., Pasceri,P., Tan-Un,K.C., Wu,X., Harper,A., Fraser,P. and Grosveld,F. (1997) Evaluation of β-globin gene therapy constructs in single copy transgenic mice. *Nucleic Acids Res.*, **25**, 1296–1302.

Forrester,W.C., Thompson,C., Elder,J.T. and Groudine,M. (1986) A developmentally stable chromatin structure in the human β-globin gene cluster. *Proc. Natl Acad. Sci. USA*, **83**, 1359–1363.

Forrester,W.C., Epner,E., Driscoll,M.C., Enver,T., Brice,M., Papayannopoulou,T. and Groudine,M. (1990) A deletion of the human β-globin locus activation region causes a major alteration in chromatin structure and replication across the entire β-globin locus. *Genes Dev.*, **4**, 1637–1649.

Gautsch,J.W. and Wilson,M.C. (1983) Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. *Nature*, **301**, 32–37.

Grosveld,F., van Assendelft,G.B., Greaves,D.R. and Kollias,G. (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell*, **51**, 975–985.

Hawley,R.G., Lieu,F.H., Fong,A.Z. and Hawley,T.S. (1994) Versatile retroviral vectors for potential use in gene therapy. *Gene Ther.*, **1**, 136–138.

Jahner,D., Stuhlmann,H., Stewart,C.L., Harbers,K., Lohler,J., Simon,I. and Jaenisch,R. (1982) *De novo* methylation and expression of retroviral genomes during mouse embryogenesis. *Nature*, **298**, 623–628.

Jones,P.L., Veenstra,G.J., Wade,P.A., Vermaak,D., Kass,S.U., Landsberger,N., Strouboulis,J. and Wolffe,A.P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nature Genet.*, **19**, 187–191.

Kempler,G., Freitag,B., Berwin,B., Nanassy,O. and Barklis,E. (1993) Characterization of the Moloney murine leukemia virus stem cell-specific repressor binding site. *Virology*, **193**, 690–699.

Leboulch,P., Huang,G.M., Humphries,R.K., Oh,Y.H., Eaves,C.J., Tuan,D.Y. and London,I.M. (1994) Mutagenesis of retroviral vectors transducing human β-globin gene and β-globin locus control region derivatives results in stable transmission of an active transcriptional structure. *EMBO J.*, **13**, 3065–3076.

Lei,H., Oh,S.P., Okano,M., Juttermann,R., Goss,K.A., Jaenisch,R. and Li,E. (1996) *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development*, **122**, 3195–3205.

Lorincz,M.C., Schubeler,D., Goeke,S.C., Walters,M., Groudine,M. and Martin,D.I. (2000) Dynamic analysis of proviral induction and *de novo* methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. *Mol. Cell. Biol.*, **20**, 842–850.

Lyko,F., Brenton,J.D., Surani,M.A. and Paro,R. (1997) An imprinting element from the mouse H19 locus functions as a silencer in *Drosophila*. *Nature Genet.*, **16**, 171–173.

McClune,S.L. and Townes,T.M. (1994) Retroviral vector sequences inhibit human β-globin gene expression in transgenic mice. *Nucleic Acids Res.*, **22**, 4477–4481.

Naldini,L., Blomer,U., Gallay,P., Ory,D., Mulligan,R., Gage,F.H., Verma,I.M. and Trono,D. (1996) *In vivo* gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science*, **272**, 263–267.

Nan,X., Ng,H.H., Johnson,C.A., Laherty,C.D., Turner,B.M., Eisenman,R.N. and Bird,A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature*, **393**, 386–389.

Niwa,O., Yokota,Y., Ishida,H. and Sugahara,T. (1983) Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell*, **32**, 1105–1113.

Okano,M., Xie,S. and Li,E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nature Genet.*, **19**, 219–220.

Okano,M., Bell,D.W., Haber,D.A. and Li,E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. *Cell*, **99**, 247–257.

Osborne,C.S., Pasceri,P., Singal,R., Sukonnik,T., Ginder,G.D. and Ellis,J. (1999) Amelioration of retroviral vector silencing in locus control region β-globin-transgenic mice and transduced F9 embryonic cells. *J. Virol.*, **73**, 5490–5496.

Pazin,M.J. and Kadonaga,J.T. (1997) What's up and down with histone deacetylation and transcription? *Cell*, **89**, 325–328.

Pirrotta,V. (1998) Polycombing the genome: PcG, trxG and chromatin silencing. *Cell*, **93**, 333–336.

Rasmussen,J.A. and Gilboa,E. (1987) Significance of DNase I-hypersensitive sites in the long terminal repeats of a Moloney murine leukemia virus vector. *J. Virol.*, **61**, 1368–1374.

Razin,A. (1998) CpG methylation, chromatin structure and gene silencing—a three-way connection. *EMBO J.*, **17**, 4905–4908.

Ryan,T.M., Behringer,R.R., Martin,N.C., Townes,T.M., Palmiter,R.D. and Brinster,R.L. (1989) A single erythroid-specific DNase 1 super-hypersensitive site activates high levels of human β-globin gene expression in transgenic mice. *Genes Dev.*, **3**, 314–323.

Schubeler,D., Francastel,C., Cimbora,D.M., Reik,A., Martin,D.I. and Groudine,M. (2000) Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human β-globin locus. *Genes Dev.*, **14**, 940–950.

Simmen,M.W., Leitgeb,S., Charlton,J., Jones,S.J., Harris,B.R., Clark,V.H. and Bird,A. (1999) Nonmethylated transposable elements and methylated genes in a chordate genome. *Science*, **283**, 1164–1167.

Stewart,C.L., Stuhlmann,H., Jahner,D. and Jaenisch,R. (1982) *De novo* methylation, expression and infectivity of retroviral genomes introduced into embryonal carcinoma cells. *Proc. Natl Acad. Sci. USA*, **79**, 4098–4102.

Strahl,B.D. and Allis,C.D. (2000) The language of covalent histone modifications. *Nature*, **403**, 41–45.

Thompson,T. and Fan,H. (1985) Mapping of DNase I-hypersensitive sites in the 5′ and 3′ long terminal repeats of integrated Moloney murine leukemia virus proviral DNA. *Mol. Cell. Biol.*, **5**, 601–609.

Tuan,D. and London,I.M. (1984) Mapping of DNase I-hypersensitive sites in the upstream DNA of human embryonic ε-globin gene in K562 leukemia cells. *Proc. Natl Acad. Sci. USA*, **81**, 2718–2722.

Tucker,K.L., Beard,C., Dausmann,J., Jackson-Grusby,L., Laird,P.W., Lei,H., Li,E. and Jaenisch,R. (1996) Germ-line passage is required for

establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.*, **10**, 1008–1020.
Tweedie,S., Ng,H.H., Barlow,A.L., Turner,B.M., Hendrich,B. and Bird,A. (1999) Vestiges of a DNA methylation system in *Drosophila melanogaster*? *Nature Genet.*, **23**, 389–390.
Urieli-Shoval,S., Gruenbaum,Y., Sedat,J. and Razin,A. (1982) The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett.*, **146**, 148–152.
Walsh,C.P. and Bestor,T.H. (1999) Cytosine methylation and mammalian development. *Genes Dev.*, **13**, 112–124.
Wei,Y., Yu,L., Bowen,J., Gorovsky,M.A. and Allis,C.D. (1999) Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. *Cell*, **97**, 99–109.
Weinberg,K.I. and Kohn,D.B. (1998) Gene therapy for congenital lymphoid immunodeficiency diseases. *Semin. Hematol.*, **35**, 354–366.
Yoder,J.A., Walsh,C.P. and Bestor,T.H. (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.*, **13**, 335–340.
Zink,D. and Paro,R. (1995) *Drosophila* Polycomb-group regulated chromatin inhibits the accessibility of a *trans*-activator to its target DNA. *EMBO J.*, **14**, 5660–5671.
Zufferey,R., Dull,T., Mandel,R.J., Bukovsky,A., Quiroz,D., Naldini,L. and Trono,D. (1998) Self-inactivating lentivirus vector for safe and efficient *in vivo* gene delivery. *J. Virol.*, **72**, 9873–9880.

Received August 2, 2000; revised September 4, 2000;
accepted September 5, 2000

TAB 14

This material may be protected by Copyright law (Title 17 U.S. Code)

Reviews in Medical Virology

Rev. Med. Virol. 2001; 11: 205–217.
DOI: 10.1002/rmv.316

REVIEW

Silencing of gene expression: implications for design of retrovirus vectors

Dylan Pannell and James Ellis*

Programs in Developmental Biology, and Cancer and Blood Research, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada, M5G1X8; and Department of Molecular and Medical Genetics, University of Toronto, Ontario, Canada

SUMMARY

Transcriptional silencing of retroviruses poses a major obstacle to their use as gene therapy vectors. Silencing is most pronounced in stem cells which are desirable targets for therapeutic gene delivery. Many vector designs combat silencing through *cis*-modifications of retroviral vector sequences. These designs include mutations of known retroviral silencer elements, addition of positive regulatory elements and insulator elements to protect the transgene from negative position effects. Similar strategies are being applied to lentiviral vectors that readily infect non-dividing quiescent stem cells. Collectively these *cis*-modifications have significantly improved vector design but optimal expression may require additional intervention to escape completely the *trans*-factors that scan for foreign DNA, establish silencing in stem cells and maintain silencing in their progeny. Cytosine methylation of CpG sites was proposed to cause retroviral silencing over 20 years ago. However, several studies provide evidence that retrovirus silencing acts through methylase-independent mechanisms. We propose an alternative silencing mechanism initiated by a speculative stem cell-specific 'somno-complex'. Further understanding of retroviral silencing mechanisms will facilitate better gene therapy vector design and raise new strategies to block transcriptional silencing in transduced stem cells. Copyright © 2001 John Wiley & Sons, Ltd.

Accepted: 20 March 2001

INTRODUCTION

Retroviruses are attractive vectors for delivery of therapeutic genes [1,2]. They have the highly desirable property of stable integration at single copy into the host genome. Often the targets for therapeutic gene delivery are stem cells, which have the ability to repopulate an entire system. However, this presents a problem as retroviruses are transcriptionally silenced in these primitive cell types. This finding has acted as a strong impetus to identify negative regulatory elements in retroviral vector sequences and to determine the underlying mechanism of retroviral gene silencing in stem cells. Many vectors with mutated silencer elements and improved positive regulatory elements have been developed to escape stem cell-specific silencing but with only partial success. Lentiviruses are now attractive alternatives to classical retroviral vectors [3], due to their ability to readily infect quiescent non-cycling stem cells [4,5]. Unfortunately, lentivirus vectors also appear to suffer from the same silencing phenomenon [6].

The cause of retrovirus vector silencing has been attributed to *de novo* cytosine methylation of CpG sequences and subsequent histone deacetylation leading to chromatin condensation [7]. However, data from several studies suggest that methylation acts only as a secondary or associated step in the retroviral silencing pathway [6,8–10].

*Corresponding author: Dr J. Ellis, Programs in Developmental Biology, and Cancer and Blood Research, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G1X8.
E-mail: jellis@sickkids.on.ca

Abbreviations used:
5-azaC, 5-azacytidine; cHS4, chicken β-globulin HS4; ddm1, decrease in DNA methylation; DR, direct repeat; EC, embryonic carcinoma; ELP, embryonic LTR binding protein; ES, embryonic stem; HDAC, histone deacetylases; HSC, haematopoietic stem cells; HSC1, deleted form of MSCV; ICF, immunodeficiency, centromere instability and facial anomalies; LCR, locus control region; MBD, methyl-CpG binding domain; MD, MPSV, dl587 rev; MEL, murine erythroleukaemia; MESV, murine embryonic stem cell virus; MND, MPSV, NCR, dl587 rev; mom1, Morpheus' molecule; MoMLV, Moloney murine leukaemia virus; MoMSV, Moloney murine sarcoma virus; MPSV, myeloproliferative sarcoma virus; MSCV, derivative of MESV; NCR, negative control region; PBS, primer binding site; PCMV, mutant form of MPSV; Pleo, pleiohomeotic; RRE, rev responsive element; SCID, severe combined immune deficiency; SIN, self-inactivating; SRC, severe combined immune deficiency repopulating cells

Copyright © 2001 John Wiley & Sons, Ltd.

Here, we discuss general mechanisms of gene silencing in eukaryotes, strategies taken to design vectors that escape silencing and models of the retroviral silencing mechanism itself.

THE PROBLEM OF RETROVIRAL SILENCING

Experiments performed some 20 years ago showed that Moloney murine leukaemia virus (MoMLV) fails to replicate after infection of pre-implantation mouse embryos or embryonic carcinoma (EC) cell lines [11–13]. Further work with murine embryonic stem (ES) cells or haematopoietic stem cells (HSCs) also found that, although the virus integrates into the DNA of the host cell, transcription from the viral LTR promoter is blocked [14]. Interestingly, retroviral silencing seems to be restricted to primitive 'stem' cell types, as MoMLV expresses normally in post-implantation embryos [12] and haematopoietic progenitors [15]. This poses a particular problem to retroviral vectors used for gene therapies. Often, gene therapy is targeted to stem cells that have the ability to repopulate an entire system. Therefore, introduction of a therapeutic gene into a stem cell can result in the gene being present in all cells of that lineage. Recent experiments with human SRCs (SCID repopulating cells, a type of HSC) confirm that retroviral silencing occurs in humans and thus will likely pose a significant problem to clinical application of retroviral mediated gene therapy [16].

CHROMATIN STRUCTURE AND GENE SILENCING

Chromatin structure modulation plays a dominant role in the control of gene expression. Therefore, a basic discussion of chromatin structure is required in order to understand possible mechanisms of retroviral silencing and strategies undertaken to overcome it. The fundamental unit of chromatin is the nucleosome, the protein component around which DNA is wrapped [17,18]. Arrays of nucleosomes form higher order chromatin structures. Each nucleosome is composed of an octamer of the histone proteins H2A, H2B, H3 and H4. In closed chromatin the linker histone, H1, is present and acts to link neighbouring nucleosomes together. There is increasing support for a modelin which higher order chromatin structure is controlled by a combinatorial code of post-translational histone modifications [19]. Histones can be acetylated, phosphorylated, ubiquitinated, ADP-ribosylated and methylated.

The best studied of these modifications is the process of acetylation and deacetylation because the enzymes that catalyse this reversible reaction have been identified and rigorously characterised [20,21]. The deacetylation of histones H3 and H4 and the presence of H1 has been correlated with silenced genes [22]. Interestingly, the histone deacetylases HDAC1 and HDAC2 have been found to be part of a multi-protein complex which interacts with the methyl-CpG binding domain (MBD) protein MeCP2 through the Sin3a co-repressor (herein referred to as a MBD/HDAC complex) [23,24]. Several other HDACs have been identified [25] but their role in methylation-dependent silencing is unknown.

The link between histone deacetylation and DNA methylation is of particular interest to retroviral silencing. Silenced proviral LTRs are heavily methylated at CpG dinucleotides by DNA methyltransferases (alternatively termed methylases) [12,13,26,27]. Moreover, DNA methylation-directed histone deacetylation suggests a model for retroviral silencing [28–30]. First, newly integrated retroviral DNA would become *de novo* methylated in stem cells. This methylated DNA would then be a target of the MBD/HDAC complex which catalyses the deacetylation of histones H3 and H4, leading to chromatin condensation and subsequent silencing of the provirus. Consistent with this idea, silenced retroviral vector sequences in transgenic mice have condensed inaccessible chromatin and are marked by a repressive histone code composed of deacetylated histone H3 and an abundance of H1 [6]. Such chromatin modifications are not limited to virus-derived sequences, but also spread to shutdown expression of a linked reporter gene.

The description of the repressive histone code on silent retrovirus vectors is still incomplete. It will be important to examine other histone modifications to determine whether they correlate with silencing, and perhaps survey modifications at specific histone residues. The complete repressive histone code may yield clues to the mechanism of retrovirus silencing. For example, two recent reports have identified mammalian homologues of the *Drosophila* gene *Su(var)3-9* as histone

Copyright © 2001 John Wiley & Sons, Ltd.

H3 specific methyltransferases, suggesting that histone methylation may be involved in gene silencing [31,32]. Further, it has been suggested that the combinatorial histone code can be interpreted by secondary factors that facilitate silencing [19]. It appears certain that chromatin modifications participate in retrovirus silencing, yet it is less clear how a stem cell senses the presence of retrovirus sequences and recruits chromatin remodelling factors to shutdown expression.

IDENTIFICATION OF RETROVIRAL SILENCER ELEMENTS

Silencer elements are DNA binding sites for *trans*-acting factors that directly or indirectly reduce transcriptional initiation at promoters. Retrovirus vectors have at least four separate silencer elements located in the LTRs and the adjacent primer binding site (PBS). *Trans*-acting factors bind to all these silencers and therefore could directly or indirectly recruit chromatin remodelling complexes. The only CpG-rich silencer is located in the LTR promoter, indicating that methylation-induced chromatin remodelling does not cause silencing mediated by the other three known silencers. Retrovirus silencer sequences were first discovered by selection for variants of MoMLV that replicated well in EC cell lines [33]. Flanagan and colleagues identified a conserved upstream region in the LTR (now called the negative control region or NCR) that reduced virus expression [34]. Ostertag's group identified mutations in the direct repeat (DR) of the LTR enhancer that increased expression in such stem cells [33]. The groups of Jaenisch and Goff found mutant PBS sites that increased expression from the LTR [35,36].

Subsequent molecular characterisation of the factors that bound to these regions identified several negative *trans*-acting factors. Specifically, the NCR binds the multi-functional transcriptional regulator YY-1 [37], six distinct nuclear factors bind to each of the direct repeats [38], and the PBS is bound by factor A [39]. In addition, the embryonic LTR-binding protein (ELP), a homologue of *Drosophila* FTZ-F1 [40], was found to bind between the NCR and DR [41]. It is interesting to note that save for ELP and factor A, all of the *trans*-acting factors that are known to bind the MoMLV LTRs are either ubiquitously expressed or not restricted to embryonic cells. As silencing still occurs when the ELP and factor A binding sites are mutated (discussed below), the stem-cell specificity of retrovirus silencing remains unclear. In addition, it has been shown that no single silencer element is sufficient to induce complete silencing [42], suggesting that silencer elements additively increase the probability of silencing.

GENERATION OF MODIFIED RETROVIRAL VECTORS

The identification and characterisation of multiple *cis*-acting sequences that negatively affect retroviral expression paved the way for the construction of numerous modified vectors that lack these sequences. Currently, many MoMLV-based vectors exist (Figure 1). PCMV, a mutant form of myeloproliferative sarcoma virus (MPSV), contains a deletion of one of the DRs, as well as a mutation in the ELP binding site and a point mutation that creates an Sp1 binding site [33]. Combination of the PCMV LTR with the dl587rev PBS which does not bind factor A, yields the

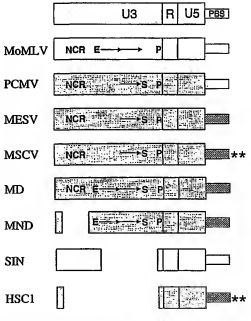


Figure 1. Schematic comparison of retroviral vectors. First construct shows general structure of retroviral vector sequence including U3, R, U5 and primer binding site (PBS). MoMLV sequences are diagrammed in white, MPSV derived sequences diagrammed in grey. E indicates ELP binding site, S indicates Sp1 binding site, P indicates CpG-rich promoter. Arrows indicate direct repeats. Hatched box indicates dl587rev PBS. ** indicates that modifications in MSCV lie outside of the vector sequence. See text for further details

Copyright © 2001 John Wiley & Sons, Ltd.

murine embryonic stem cell virus (MESV) [43]. Further refinement of the MESV vector was made by Hawley *et al.* to increase titres by including an extended packaging signal from the LN vector and an upstream gag sequence from Moloney murine sarcoma virus, (MoMSV) resulting in MSCV [44]. It should be noted that MSCV retains the NCR, one of the DRs, and the CpG rich promoter. Although MSCV does offer an improvement over the wild-type virus, later studies have shown that it retains significant silencing activity [6].

Others have focused on the substitution of the MoMLV enhancer with the MPSV enhancer, which had been shown to have increased expression in EC cells [33]. Substitution of the MoMLV LTR with the MPSV LTR in conjunction with the d1587rev PBS yielded the MD (MPSV, d1587rev) vector [45]. Unfortunately, this double mutation did not show significantly increased expression. However, deletion of the NCR from the MD vector gave rise to the MND (MPSV, NCR, d1587rev) vector, which did show substantial improvement with expression in 43% of infected EC cells [45]. The MND vector also showed significantly increased expression in murine HSCs [46].

Finally, deletions which yield self-inactivating or SIN viruses have steadily gained popularity due to the decreased probability of these viruses undergoing recombination leading to the generation of a replication competent retrovirus. The first such SIN retroviruses were MoMLV-based vectors with a 299bp deletion in the 3'LTR that removes most of the DR and the promoter, but retains the NCR and wild-type PBS [47]. More recently the HSC1 vector was constructed by deletion of the entire NCR, DR, and part of the CpG rich promoter sequences in MSCV which retains the mutant PBS. This SIN vector directs expression in 71% of transduced EC cells, offering a further improvement still [42]. The HSC1 vector is particularly attractive both for its ability to express in non-permissive environments and its SIN properties.

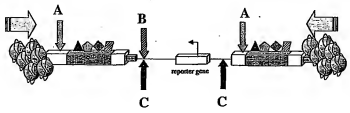
In summary, progressive deletion or mutation of *cis*-acting silencer sequences has led to the development of retroviral vectors that have an increased probability of expression. Nevertheless, no current vector is able to express at all integration sites despite mutations in all known retroviral silencer elements, suggesting that novel silencer sequences remain to be discovered.

POSITIVE REGULATORY ELEMENTS AND INSULATORS

Other attempts to increase expression from retrovirus transduced genes have focused on the addition of positive regulatory elements to the expression cassette. One example of these experiments incorporates the β-globin locus control region (LCR). The LCR directs position-independent, copy number-dependent expression in transgenic mice, and is therefore a potent activator of transcription [48,49]. Notably, several experiments show that MoMLV and MSCV retrovirus vector sequences dominantly silence LCR/β-globin expression cassettes in transgenic mice [6,50]. These data underscore the potency of the silencing effect exerted by retroviral vector sequences. As some of the sequences contained in the β-globin LCR are deleterious to retroviral replication and therefore difficult to transmit [51,52], other groups have focused on using heterologous promoters to achieve higher levels of expression [53–55]. Although these promoters are transmitted more efficiently, they will likely still be subject to the dominant silencing activity of the vector sequences.

A different strategy to obtain high levels of expression by modification of the vector is to include insulator elements. Insulators are believed to form expression boundaries [56] and are assayed by their ability to block positive and negative position effects when they flank a reporter gene [57]. Insulators located between an enhancer and a promoter also block transcriptional activation [58,59]. Their ability to block retroviral silencers is under investigation. Encouraging preliminary results were obtained when a monomer of the chicken β-globin HS4 (cHS4) insulator was inserted into the 3' LTR of MoMLV vectors (Figure 2 position A) [60,61]. In these experiments cHS4 blocked negative position effects in mature cell types, but by virtue of its placement in the U3 region of the LTR was unable to block retroviral silencers that were more proximal to the reporter gene in ES cells. Similar experiments with the cHS4 insulator placed between the 5'LTR and the reporter gene failed to protect from silencing originating at the 3'LTR (Figure 2 position B) [62]. These experiments underscore the need for insulators to be placed flanking the internal gene in such a way that they shield it from position effects and both LTRs

Copyright © 2001 John Wiley & Sons, Ltd.



	Insulator Position		
	A	B	C
Position Effect Block	+	-	+
Retroviral Silencing Block	-	-	+

Figure 2. Placement of insulator elements in retrovirus vectors. Integrated retroviruses are subject to silencing both from position effects (symbolised by spreading nucleosomes) and silencer elements located within the LTRs (grey regions bound by polygons). Insertion of insulators into position (A) shields reporter from position effects but fails to block silencers located in the 5'LTR. Insertion into position B blocks neither 3'position effects nor silencers in the 3'LTR. Only insertion into flanking position C completely shields from both position effects and retroviral silencer elements

(Figure 2 position C). In fact, retrovirus silencing is completely relieved by such flanking arrangements of the cHS4 insulator, permitting an LCRβ-globin reporter transgene to be expressed in a copy number-dependent, position-independent manner in mice (C. S. Osborne, P. Pasceri and JE, unpublished results). It is not clear whether this shielding is sufficient to confer position-independent activation on commonly used promoters such as PGK in the absence of an LCR.

LENTIVIRAL VECTORS

The newest and perhaps most promising vectors currently being developed for the infection of non-dividing quiescent stem cells are HIV-1 based lentiviral vectors. These vectors do not rely on the passive mechanism of nuclear envelope breakdown during the cell cycle to gain entry to the nucleus. Instead they utilise active transport mechanisms and achieve entry through the nucleopore [63]. This feature is particularly relevant when infecting slowly dividing stem cells, because a greater transduction efficiency is achieved in comparison to MoMLV-based retroviruses [4]. Another attractive property of lentiviral vectors is their ability to stabilise the RNA genome during replication by virtue of the rev responsive element (RRE). In what has become a benchmark experiment in β-globin gene therapy, an HIV-1 based vector called HR' was used to transfer efficiently an LCRβ-globin expression cassette into bone marrow stem cells from β-thalassaemic mice [64]. The level of protein expression of the transduced human β-globin gene was approximately 13%, sufficient to correct the anaemic phenotype. Further experiments with lentiviruses have shown that they are able to transduce efficiently human SRCs [65,66], which will be necessary for extension of the findings in mice to clinical trials.

However, it is evident that lentiviruses are subject to the same potent silencing effects as MoMLV-based retroviruses [67,68]. HIV-1 sequences have been shown to dominantly silence a linked LCRβ-globin reporter gene in transgenic mice [6]. Fortunately, the findings in MoMLV-based retroviruses seem to translate well to their lentiviral cousins. A SIN version of the HR' vector is available (SIN-18) which has a large deletion of the U3 region similar to HSC1 [69]. This deletion has been shown not to interfere with viral titres, and to partially relieve silencing of a linked LCRβ-globin gene in transgenic mice [6]. These SIN-18 vectors hold much promise, although further improvements are necessary before they become widely utilised for therapeutic gene transfer.

MECHANISMS OF RETROVIRAL SILENCING

Despite many years of study, the cause of the stem cell-specific block to transcription of retroviruses remains unknown. The previous discussion has concentrated on the finding that multiple

Copyright © 2001 John Wiley & Sons, Ltd.

mutations in *cis*-acting sequences provide partial relief of silencing, but may not abrogate it completely. Developing a better understanding of the mechanism of retroviral silencing may lead to strategies to overcome or evade its establishment.

THE GENOME DEFENCE HYPOTHESIS

The genome defence hypothesis contends that cytosine methylation arose in mammals to restrict expression of intragenomic parasites [28]. Evidence supporting the genome defence hypothesis stems from the finding that the LTRs of retrotransposable elements such as L1s, IAPs and endogenous retroviruses are all heavily methylated at CpG dinucleotides [7]. Similarly, the methylation inhibiting drug 5-azacytidine (5-azaC) has been used to reactivate silent proviruses in EC cells [13]. However, careful inspection of the data reveals that reactivation induced by 5-azaC only increases the retroviral titre to 0.05% of a non-silenced control. This suggests that other methylation-independent mechanisms are actively silencing the majority of retroviruses in these cells. Further evidence supporting a causal role for methylation in retroviral silencing comes from the finding that the IAP class of endogenous retrovirus becomes reactivated in ES cells that are null for the maintenance methyltransferase, *Dnmt*1 [70]. *Dnmt*1 is expressed throughout development [71] and has a preference for hemimethylated DNA [26]. Importantly, C-type retroviruses including MoMLV remain methylated and transcriptionally silent in these *Dnmt*1 null ES cells [72]. Therefore, it is not possible to determine the role of methylation in silencing of C-type retroviruses using these cells.

The continued ability of *Dnmt*1 null ES cells to methylate exogenous retroviral LTRs is due to the existence of two *de novo* methyltransferases, *Dnmt*3a and *Dnmt*3b [26]. In contrast to *Dnmt*1, the *de novo* methyltransferases act on both hemimethylated and unmethylated DNA [26]. Further, their pattern of expression correlates well with retroviral silencing, as they are expressed to high levels in ES cells and HSCs [26,73]. This may be seen as a further link between methylation and silencing. In this model, viral infection of cells that express one or both of the *de novo* methylases (i.e. ES cells and HSCs) would lead to methylation of the integrated provirus. These methyl-CpGs would then recruit a MBD/HDAC complex which

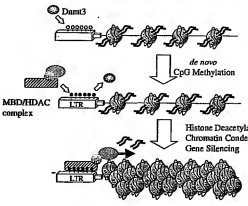


Figure 3. The genome defence hypothesis. CpGs (unfilled circles) of a newly integrated provirus inserted into open chromatin (acetylated histones) are targeted for methylation by a *de novo* methylase. Methylated CpGs (filled circles) are bound by MBD/HDAC complex. Deacetylation of histones results in chromatin condensation and gene silencing

catalyses histone deacetylation leading to chromatin condensation and gene silencing (Figure 3). These complexes are ubiquitously expressed and therefore would act to propagate the silenced state throughout differentiation. The ubiquitously expressed *Dnmt*1 would maintain the methylated state of the silenced provirus, ensuring that it does not become reactivated during replication or differentiation.

CYTOSINE METHYLATION AND RETROVIRAL SILENCING: CAUSE OR CONSEQUENCE?

The first studies to cast doubt on whether methylation causes retroviral silencing examined the timing of silencing and methylation. Silencing occurs in EC cells 2 days post-infection whereas methylation is not detectable until 8–16 days post infection [8,9]. Pre-infection differentiation caused the cells to be permissive for retroviral expression. In contrast, differentiation of cells after infection did not cause the retroviruses to become reactivated. These data suggest that the factors required for establishment of silencing are stem cell-specific, but that those required for maintenance of the silenced state are ubiquitous. Interestingly, the methylation inhibiting drug 5-azaC was able to reactivate expression in differentiated cells, whereas it failed to reactivate retroviral expression in undifferentiated EC cells. These results can be interpreted as evidence that methylation is

Copyright © 2001 John Wiley & Sons, Ltd.

required for the maintenance of silencing of retroviruses, but is dispensable for its establishment.

Transcriptional extinction of retrovirally transduced reporter genes may use similar silencing mechanisms. In these experiments, MoMLV vector infected murine erythroleukaemia (MEL) cells which initially expressed were sorted and silencing was examined over time. It was found that an initial extinction event is reversible by the HDAC inhibitor trichostatin A. At later time points, reactivation of expression required 5-azaC in conjunction with trichostatin A, suggesting that methylation is a secondary step associated with maintenance of silencing [74]. It is important to note that although extinction and silencing are conceptually related phenomena, extinction experiments focus on a population of cells that initially express. In contrast, silencing refers to integration events in undifferentiated stem cells that are immediately shut off. In spite of these differences, it appears that both processes employ an initial methylation-independent silencing event likely to involve chromatin remodelling, with methylation being a secondary or associated step.

METHYLASE INDEPENDENT RETROVIRAL SILENCING

The debate surrounding the role of methylation in retroviral silencing has been confounded by the absence of a methylase-free environment to test the causal role of methylation. This is due to the multiple methyltransferases present in murine cells. As previously mentioned, the finding that exogenous retroviral LTRs become methylated in *Dnmt1* knock-out ES cells demonstrated the existence of other methylases. A second methyltransferase, *Dnmt2* was identified based on homology of its catalytic domain to that of other methyltransferases [75]. Somewhat frustratingly, ES cells null for *Dnmt2* also retain the ability to methylate exogenous MoMLV LTRs. Further, mice generated from *Dnmt2* homozygous null ES cells develop normally, in sharp contrast to the lethal phenotype observed in *Dnmt1*-/- mice [76]. Biochemical assays failed to detect any methyltransferase activity from *Dnmt2 in vitro*. To date it is not clear what function *Dnmt2* serves.

In an attempt to bypass the complex situation in mammals, several recent studies have examined the behaviour of mammalian silencers in *Drosophila* [77,78], which does not rely on methylation for maintenance of a repressed state. In fact, until recently, the *Drosophila* genome was thought to be devoid of any methylation [79], and it has no genes encoding known cytosine methylases. When assayed in *Drosophila*, retroviral and lentiviral vector sequences exhibit strong silencing activity on a linked UAS*lacZ* reporter transgene [6]. These data demonstrate that retroviral silencing is a conserved function that can occur by methylase-independent mechanisms. However, two recent reports have detected trace amounts of methylcytosine in *Drosophila*, which may [80] or may not [81] be restricted to the early embryo. Whether this low-level methylation has any biological significance has yet to be tested. A *Drosophila* homologue of the vertebrate methyl-CpG binding domain proteins is known to exist, named dMBD2/3 for its shared homology with both MBD2 and MBD3 [82]. This protein interacts with a *Drosophila* HDAC complex both *in vitro* and *in vivo*. Curiously, it has lost all ability to bind methylated CpG dinucleotides, indicating that dMBD2/3 is conserved for HDAC dependent silencing alone.

An end to this lengthy debate was finally in sight with the identification and cloning of the long sought-after *de novo* methylases, *Dnmt3a* and *Dnmt3b* [26]. Both of these genes are expressed to high levels in ES cells and *de novo* methylate retroviral LTRs *in vitro*. Homozygous knock-out of each of these genes is lethal in mice, demonstrating the essential role methylation must play in normal mammalian development [27]. When exogenous retroviral LTRs were examined for methylation in ES cells null for either *Dnmt3a* or *Dnmt3b* alone it was found that the retroviral sequence was still methylated, suggesting overlapping functions of these two genes in this process. However, in ES cells doubly null for both *Dnmt3a* and *Dnmt3b*, exogenous retroviral LTRs remain unmethylated.

Using these double knock-out ES cells it was possible to determine whether *de novo* methylation is required for establishment of retroviral silencing. Infection of *Dnmt3a-/-; Dnmt3b-/-* ES cells with either MoMLV, MSCV or HSC1 based vectors with internal PGK promoter eGFP reporter cassettes showed no observable relief of silencing compared with the wild-type parental ES cell line [6]. These experiments demonstrate that *de novo*

Copyright © 2001 John Wiley & Sons, Ltd.

methylases are not required for establishment of retroviral silencing in mammalian stem cells. Further, it clearly demonstrates that other, as yet unidentified factors, must exist which establish the silenced state.

TOWARDS AN INTEGRATED MODEL OF RETROVIRAL SILENCING

What then is the role of DNA methylation? The embryonic lethality displayed by mice which are null for *Dnmt*1, *Dnmt*3a or *Dnmt*3b emphatically demonstrates that it is an indispensable function, essential for normal development. Similarly, mutations in *DNMT*3B lead to immunodeficiency, centromere instability and facial anomalies (ICF) syndrome in humans [27,83,84], reinforcing its importance. We suggest that methylation acts to maintain an epigenetic signal initiated by stem cell-specific silencing factors. The covalent modification of adding a methyl group to a cytosine residue is mitotically heritable, and thus suitable for silencing genes in the long term. In this model, factors that are restricted to stem cells are responsible for the initial silencing event, likely through chromatin remodelling. By extension, these same factors may be responsible for marking the silenced sequence as a target for a *de novo* methylase.

We refer collectively to these proposed stem cell-specific factors as a 'somno-complex' for its ability to induce a dormant state of gene inactivity prior to long-term silencing by methylation (Figure 4). We speculate that the exogenous retrovirus in stem cells is detected by the putative somno-complex and LTR transcription is inactivated. This model is consistent with all observations relating to retroviral silencing mechanisms. It explains the ability of 5-azaC to reactivate some expression in differentiated cells that lack the somno-complex, but not undifferentiated EC cells that retain the complex. It also clarifies why retroviral silencing is observed prior to methylation and why it can occur independently of *de novo* methylases in undifferentiated ES cells.

We can only speculate on the factors that make up the somno-complex. In all likelihood, they would include stem cell-specific chromatin remodelling factors. If this complex is indeed required to direct the *de novo* methylases to retroviral LTRs, then they would necessarily share the same

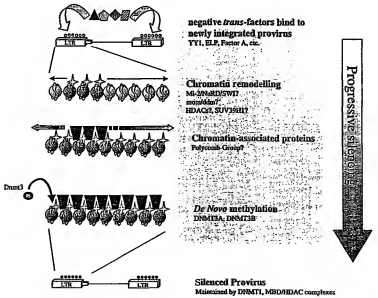


Figure 4. Speculative somno-complex model of retroviral silencing. Newly integrated provirus is bound by negative *trans*-acting factors. Somno-complex mediated steps occur in stem cells and act to silence transcription. Chromatin over the provirus is remodelled (darker nucleosomes) and modifications (stars) such as acetyl groups or methyl groups are attached to histones. These modifications may be recognised as a repressive histone code by chromatin-associated proteins such as members of the *Polycomb*-group. Targeting of *de novo* methylase to these sites allows methylation of the retroviral LTR, resulting in the propagation of a heritable silenced state in differentiated cells that lack a functional somno-complex

Copyright © 2001 John Wiley & Sons, Ltd.

expression patterns as *de novo* methylases. Recently, detailed analysis of genes upregulated in murine haematopoietic stem cells yielded several factors that are candidates, including a novel HDAC and an HSC-specific *Polycomb*-group gene (*Pc*-G). The *de novo* methylase *Dnmt*3b is also expressed to high levels in murine HSCs as expected [73].

Possible somno-complex members could be assembled on the NCR silencer of retroviral LTRs. The NCR is bound by the YY-1 factor [37], which in turn is known to interact with several members of the HDAC family [85,86]. In stem cells, YY-1 bound to the NCR could be recognised by a stem cell-specific HDAC. Alternatively, the *Drosophila* homologue of YY-1 is the *pleiohomeotic* gene (*pho*) [87]. *Pho* is the only known *Pc*-G gene that specifically binds DNA, and it does this at consensus YY-1 binding sites. *Pho* is believed to tether *Pc*G silencer complexes to target DNA. By analogy, YY-1 in mammalian stem cells may tether a stem cell-specific *Pc*G protein complex onto retroviral LTRs. *Pc*-G complexes in flies may also repress transcription through the dMi-2 component of another HDAC complex. This complex is known as NuRD, is conserved in mammals and also contains MBD3 (for review see [88]). Regardless of whether an HDAC or *Pc*G protein is recruited by YY-1, it is important to remember that retrovirus silencers act additively [45], and that the NCR alone has little silencing activity [42]. Thus the somno-complex may not be fully competent until multiple components are recruited to multiple silencers.

Perhaps the most intriguing possibility for somno-complex members comes from experiments on transgene silencing in the plant *Arabidopsis thaliana* [89,90]. Two separate mutagenesis screens were performed to identify factors required for silencing of a reporter gene. The first gene to be found was *ddm*1 (for decrease in DNA methylation), mutations in which relieved transgene silencing concomitant with a decrease in methylation [91]. As is expected for mutations that abrogate methylation, these plants showed pleiotropic developmental defects. What is particularly interesting with regard to the proposed somno-complex is that *ddm*1 shows no discernible homology to methyltransferases. Rather, it is related to the SWI2/SNF2 family of ATP-dependent chromatin remodelling helicases [92,93]. To explain methylation defects, *ddm*1 would have to modulate or direct the activity of the methylase as predicted for the somno-complex. Whether it does this directly or indirectly is not currently known.

A second screen, performed by Amedeo *et al.*, identified another plant mutant called *mom*1 (for Morpheus' molecule) that also relieves transgene silencing [94]. *Mom*1 mutants do not show any defects in methylation, and the protein does not show any homology to MBD proteins. To the contrary, transgenes remain methylated in *mom*1 mutants but express normally. Like *ddm*1, the *mom*1 protein also resembles a SWI2/SNF2 helicase. This uncoupling of methylation and silencing provides further evidence that they are separable processes, and are likely to be parallel pathways that share some chromatin remodelling functions. Unlike *ddm*1, *mom*1 mutants develop normally, arguing against a role in methylation mediated silencing. If mammalian *mom*1 homologues exist, it is unclear how they would be recruited to target DNA sequences. Nevertheless, it seems fitting that a molecule named for the Greek god of dreams is an excellent candidate member of the putative somno-complex that silences retrovirus vectors in stem cells.

FUTURE PROSPECTS

The somno-complex concept merits testing by searching for stem cell-specific components involved in the establishment of retrovirus vector silencing. Identification of the unknown factors required for initiation of retroviral silencing in stem cells, coupled with a better understanding of the repressive histone code and the role of cytosine methylation in its maintenance will increase our understanding of gene silencing in general. Despite stem cell-specific silencing, retroviruses remain one of the most popular and best-suited vectors for gene therapy. Self-inactivating vectors with multiple modifications of negative *cis*-acting binding sites are powerful tools to achieve this end. Further vector improvements will use powerful LCR elements or shield therapeutic genes with flanking internal insulator elements, and self-inactivating lentiviral vectors may direct safe and efficient gene transfer into quiescent stem cells. Administration of HDAC-inhibitors in concert with methylation-inhibiting drugs may help to prevent extinction of retrovirally transduced genes in mature cells. Nevertheless, in stem cells where both establishment

Copyright © 2001 John Wiley & Sons, Ltd.

factors and methylases are present, it is clearly not sufficient to inhibit methylation alone. Ultimately, identification of the stem cell-specific factors that are responsible for initiation of retroviral silencing may suggest alternative interventions to inhibit this process before it is established.

ACKNOWLEDGEMENTS

We acknowledge the Canadian Institutes for Health Research (CIHR) grant support (to JE). DP is supported by a CIHR Doctoral Research Award. We regret the omission of many fine studies due to space constraints.

REFERENCES

1. Glorioso JC, Naldini L, Kay MA. Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics. *Nat Med* 2001; 7: 33–40.
2. Palu G, Parolin C, Takeuchi Y, Pizzato M. Progress with retroviral gene vectors. *Rev Med Virol* 2000; 10: 185–202.
3. Vigna E, Naldini L. Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. *J Gene Med* 2000; 2: 308–316.
4. Uchida N, Sutton RE, Friera AM, *et al.* HIV, but not murine leukemia virus, vectors mediate high efficiency gene transfer into freshly isolated G0/G1 human hematopoietic stem cells. *Proc Natl Acad Sci U S A* 1998; 95: 11939–11944.
5. Sutton RE, Wu HT, Rigg R, Bohnlein E, Brown PO. Human immunodeficiency virus type 1 vectors efficiently transduce human hematopoietic stem cells. *J Virol* 1998; 72: 5781–5788.
6. Pannell D, Osborne CS, Yao S, *et al.* Retrovirus vector silencing is *de novo* methylase independent and marked by a repressive histone code. *Embo J* 2000; 19: 5884–5894.
7. Yoder JA, Walsh CP, Bestor TH. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* 1997; 13: 335–340.
8. Gautsch JW, Wilson MC. Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. *Nature* 1983; 301(5895): 32–37.
9. Niwa O, Yokota Y, Ishida H, Sugahara T. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell* 1983; 32: 1105–1113.
10. Kempler G, Freitag B, Berwin B, Nanassy O, Barklis E. Characterization of the Moloney murine leukemia virus stem cell-specific repressor binding site. *Virology* 1993; 193: 690–699.
11. Teich NM, Weiss RA, Martin GR, Lowy DR. Virus infection of murine teratocarcinoma stem cell lines. *Cell* 1977; 12: 973–982.
12. Jahner D, Stuhlmann H, Stewart CL, *et al. De novo* methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* 1982; 298(5875): 623–628.
13. Stewart CL, Stuhlmann H, Jahner D, Jaenisch R. *De novo* methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. *Proc Natl Acad Sci U S A* 1982; 79: 4098–4102.
14. Challita PM, Kohn DB. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*. *Proc Natl Acad Sci U S A* 1994; 91: 2567–2571.
15. Klug CA, Cheshier S, Weissman IL. Inactivation of a GFP retrovirus occurs at multiple levels in long-term repopulating stem cells and their differentiated progeny. *Blood* 2000; 96: 894–901.
16. Guenechéa G, Gan OI, Dorrell C, Dick JE. Distinct classes of human stem cells that differ in proliferative and self-renewal potential. *Nature Immunol* 2001; 2: 75–82.
17. Felsenfeld G. Chromatin unfolds. *Cell* 1996; 86: 13–19.
18. Bird AP, Wolffe AP. Methylation-induced repression – belts, braces, and chromatin. *Cell* 1999; 99: 451–454.
19. Strahl BD, Allis CD. The language of covalent histone modifications. *Nature* 2000; 403(6765): 41–45.
20. Ng HH, Bird A. Histone deacetylases: silencers for hire. *Trends Biochem Sci* 2000; 25: 121–126.
21. Kuo MH, Allis CD. Roles of histone acetyltransferases and deacetylases in gene regulation. *Bioessays* 1998; 20: 615–626.
22. Schubeler D, Lorincz MC, Cimbora DM, *et al.* Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Mol Cell Biol* 2000; 20: 9103–9112.
23. Nan X, Campoy FJ, Bird A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* 1997; 88: 471–481.
24. Jones PL, Veenstra GJ, Wade PA, *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 1998; 19: 187–191.
25. Gray SG, Ekstrom TJ. The human histone deacetylase Family. *Exp Cell Res* 2001; 262: 75–83.
26. Okano M, Xie S, Li E. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 1998; 19: 219–220.
27. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are

Copyright © 2001 John Wiley & Sons, Ltd.

essential for *de novo* methylation and mammalian development. *Cell* 1999; **99**: 247–257.
28. Walsh CP, Bestor TH. Cytosine methylation and mammalian development. *Genes Develop* 1999; **13**: 112–124.
29. Eden S, Hashimshony T, Keshet I, Cedar H. DNA methylation models histone deacetylation. *Nature* 1998; 394(6696): 842.
30. Razin A. CpG methylation, chromatin structure and gene silencing – a three-way connection. *Embo J* 1998; **17**: 4905–4908.
31. Rea S, Eisenhaber F, O'Carroll D, *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* 2000; 406(6796): 593–599.
32. O'Carroll D, Scherthan H, Peters AH, *et al.* Isolation and characterization of suv39h2, a second histone H3 methyltransferase gene that displays testis-specific expression. *Mol Cell Biol* 2000; **20**: 9423–9433.
33. Hilberg F, Stocking C, Ostertag W, Grez M. Functional analysis of a retroviral host-range mutant: altered long terminal repeat sequences allow expression in embryonal carcinoma cells. *Proc Natl Acad Sci U S A* 1987; 84: 5232–5236.
34. Flanagan JR, Krieg AM, Max EE, Khan AS. Negative control region at the 5′ end of murine leukemia virus long terminal repeats. *Mol Cell Biol* 1989; **9**: 739–746.
35. Barklis E, Mulligan RC, Jaenisch R. Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. *Cell* 1986; **47**: 391–399.
36. Colicelli J, Goff SP. Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. *J Virol* 1986; **57**: 37–45.
37. Flanagan JR, Becker KG, Ennist DL, *et al.* Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. *Mol Cell Biol* 1992; **12**: 38–44.
38. Speck NA, Baltimore D. Six distinct nuclear factors interact with the 75-base-pair repeat of the Moloney murine leukemia virus enhancer. *Mol Cell Biol* 1987; **7**: 1101–1110.
39. Petersen R, Kempler G, Barklis E. A stem cell-specific silencer in the primer-binding site of a retrovirus. *Mol Cell Biol* 1991; **11**: 1214–1221.
40. Tsukiyama T, Ueda H, Hirose S, Niwa O. Embryonal long terminal repeat-binding protein is a murine homolog of FTZ-F1, a member of the steroid receptor superfamily. *Mol Cell Biol* 1992; **12**: 1286–1291.
41. Tsukiyama T, Niwa O, Yokoro K. Mechanism of suppression of the long terminal repeat of Moloney leukemia virus in mouse embryonal carcinoma cells. *Mol Cell Biol* 1989; **9**: 4670–4676.
42. Osborne CS, Pasceri P, Singal R, Sukonnik T, Ginder GD, Ellis J. Amelioration of retroviral vector silencing in locus control region beta-globin-transgenic mice and transduced F9 embryonic cells. *J Virol* 1999; **73**: 5490–5496.
43. Grez M, Akgun E, Hilberg F, Ostertag W. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. *Proc Natl Acad Sci U S A* 1990; **87**: 9202–9206.
44. Hawley RG, Lieu FH, Fong AZ, Hawley TS. Versatile retroviral vectors for potential use in gene therapy. *Gene Therapy* 1994; **1**: 136–138.
45. Challita PM, Skelton D, el-Khoueiry A, Yu XJ, Weinberg K, Kohn DB. Multiple modifications in cis elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells. *J Virol* 1995; 69: 748–755.
46. Robbins PB, Skelton DC, Yu XJ, Halene S, Leonard EH, Kohn DB. Consistent, persistent expression from modified retroviral vectors in murine hematopoietic stem cells. *Proc Natl Acad Sci U S A* 1998; **95**: 10182–10187.
47. Yu SF, von Ruden T, Kantoff PW, *et al.* Self-inactivating retroviral vectors designed for transfer of whole genes into mammalian cells. *Proc Natl Acad Sci U S A* 1986; 83: 3194–3198.
48. Grosveld F, van Assendelft GB, Greaves DR, Kollias G. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell* 1987; **51**: 975–985.
49. Ryan TM, Behringer RR, Townes TM, Palmiter RD, Brinster RL. High-level erythroid expression of human alpha-globin genes in transgenic mice. *Proc Natl Acad Sci U S A* 1989; 86: 37–41.
50. McCune SL, Townes TM. Retroviral vector sequences inhibit human beta-globin gene expression in transgenic mice. *Nucleic Acids Res* 1994; **22**: 4477–4481.
51. Sadelain M, Wang CH, Antoniou M, Grosveld F, Mulligan RC. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. *Proc Natl Acad Sci U S A* 1995; **92**: 6728–6732.
52. Leboulch P, Huang GM, Humphries RK, *et al.* Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an active transcriptional structure. *EMBO J* 1994; **13**: 3065–3076.
53. Ramezani A, Hawley TS, Hawley RG. Lentiviral vectors for enhanced gene expression in human hematopoietic cells. *Mol Ther* 2000; **2**: 458–469.
54. Sabatino DE, Cline AP, Gallagher PG, *et al.* Substitution of the human beta-spectrin promoter for the human agamma-globin promoter prevents

Copyright © 2001 John Wiley & Sons, Ltd.

silencing of a linked human beta-globin gene in transgenic mice. *Mol Cell Biol* 1998; 18: 6634–6640.
55. Sabatino DE, Wong C, Cline AP, *et al.* A minimal ankyrin promoter linked to a human gamma-globin gene demonstrates erythroid specific copy number dependent expression with minimal position or enhancer dependence in transgenic mice. *J Biol Chem* 2000; 275: 28549–28554.
56. Sun FL, Elgin SC. Putting boundaries on silence. *Cell* 1999; 99: 459–462.
57. Kellum R, Schedl P. A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 1991; 64: 941–950.
58. Kellum R, Schedl P. A group of scs elements function as domain boundaries in an enhancer-blocking assay. *Mol Cell Biol* 1992; 12: 2424–2431.
59. Chung JH, Whiteley M, Felsenfeld G. A 5′ element of the chicken beta-globin domain serves as an insulator in human erythroid cells and protects against position effect in *Drosophila*. *Cell* 1993; 74: 505–514.
60. Rivella S, Callegari JA, May C, Tan CW, Sadelain M. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. *J Virol* 2000; 74: 4679–4687.
61. Emery DW, Yannaki E, Tubb J, Stamatoyannopoulos G. A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci U S A* 2000; 97: 9150–9155.
62. Modin C, Pedersen FS, Duch M. Lack of shielding of primer binding site silencer-mediated repression of an internal promoter in a retrovirus vector by the putative insulators scs, BEAD-1, and HS4. *J Virol* 2000; 74: 11697–11707.
63. Chicurel M. Virology. Probing HIV's elusive activities within the host cell. *Science* 2000; 290: 1876–1879.
64. May C, Rivella S, Callegari J, *et al.* Therapeutic haemoglobin synthesis in beta-thalassaemic mice expressing lentivirus-encoded human beta-globin. *Nature* 2000; 406: 82–86.
65. Schilz AJ, Brouns G, Knoss H, *et al.* High efficiency gene transfer to human hematopoietic SCID-repopulating cells under serum-free conditions. *Blood* 1998; 92: 3163–3171.
66. Woods NB, Fahlman C, Mikkola H, *et al.* Lentiviral gene transfer into primary and secondary NOD/SCID repopulating cells. *Blood* 2000; 96: 3725–3733.
67. Miyoshi H, Smith KA, Mosier DE, Verma IM, Torbett BE. Transduction of human CD34+ cells that mediate long-term engraftment of NOD/SCID mice by HIV vectors. *Science* 1999; 283: 682–686.
68. Mikkola H, Woods NB, Sjogren M, *et al.* Lentivirus gene transfer in murine hematopoietic progenitor cells is compromised by a delay in proviral integration and results in transduction mosaicism and heterogeneous gene expression in progeny cells. *J Virol* 2000; 74: 11911–11918.
69. Zufferey R, Dull T, Mandel RJ, *et al.* Self-inactivating lentivirus vector for safe and efficient *in vivo* gene delivery. *J Virol* 1998; 72: 9873–9880.
70. Walsh CP, Chaillet JR, Bestor TH. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 1998; 20: 116–117.
71. Bestor T, Laudano A, Mattaliano R, Ingram V. Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J Mol Biol* 1988; 203: 971–983.
72. Lei H, Oh SP, Okano M, *et al. De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 1996; 122: 3195–3205.
73. Phillips RL, Ernst RE, Brunk B, *et al.* The genetic program of hematopoietic stem cells. *Science* 2000; 288: 1635–1640.
74. Lorincz MC, Schubeler D, Goeke SC, Walters M, Groudine M, Martin DI. Dynamic analysis of proviral induction and *de novo* methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. *Mol Cell Biol* 2000; 20: 842–850.
75. Okano M, Xie S, Li E. Dnmt2 is not required for *de novo* and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res* 1998; 26: 2536–2540.
76. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 1992; 69: 915–926.
77. Lyko F, Buiting K, Horsthemke B, Paro R. Identification of a silencing element in the human 15q11-q13 imprinting center by using transgenic *Drosophila*. *Proc Natl Acad Sci U S A* 1998; 95: 1698–1702.
78. Lyko F, Brenton JD, Surani MA, Paro R. An imprinting element from the mouse H19 locus functions as a silencer in *Drosophila*. *Nat Genet* 1997; 16: 171–173.
79. Urieli-Shoval S, Gruenbaum Y, Sedat J, Razin A. The absence of detectable methylated bases in *Drosophila melanogaster* DNA. *FEBS Lett* 1982; 146: 148–152.
80. Gowher H, Leismann O, Jeltsch A. DNA of *Drosophila melanogaster* contains 5-methylcytosine. *Embo J* 2000; 19: 6918–6923.
81. Lyko F, Ramsahoye BH, Jaenisch R. DNA methylation in *Drosophila melanogaster*. *Nature* 2000; 408: 538–540.
82. Tweedie S, Ng HH, Barlow AL, Turner BM, Hendrich B, Bird A. Vestiges of a DNA methylation

Copyright © 2001 John Wiley & Sons, Ltd.

system in *Drosophila melanogaster*? *Nat Genet* 1999; 23: 389–390.
83. Hansen RS, Wijmenga C, Luo P, *et al.* The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* 1999; 96: 14412–14417.
84. Xu GL, Bestor TH, Bourc'his D, *et al.* Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 1999; 402: 187–191.
85. Yang WM, Inouye C, Zeng Y, Bearss D, Seto E. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. *Proc Natl Acad Sci U S A* 1996; 93: 12845–12850.
86. Thomas MJ, Seto E. Unlocking the mechanisms of transcription factor YY1: are chromatin modifying enzymes the key? *Gene* 1999; 236: 197–208.
87. Brown JL, Mucci D, Whiteley M, Dirksen M-L, Kassis JA. The *Drosophila* polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol Cell* 1998; 1: 1057–1064.
88. Ahringer J. NuRD and SIN3 histone deacetylase complexes in development. *Trends Genet* 2000; 16: 351–356.
89. Tweedie S, Bird A. Gene expression. Mutant weed breaks silence. *Nature* 2000; 405: 137–138.
90. Chandler VL, Jorgensen R. Silencing Morpheus awakens transgenes. *Nat Biotechnol* 2000; 18: 602–603.
91. Vongs A, Kakutani T, Martienssen RA, Richards EJ. *Arabidopsis thaliana* DNA methylation mutants. *Science* 1993; 260: 1926–1928.
92. Jeddeloh JA, Bender J, Richards EJ. The DNA methylation locus DDM1 is required for maintenance of gene silencing in *Arabidopsis*. *Genes Dev* 1998; 12: 1714–1725.
93. Jeddeloh JA, Stokes TL, Richards EJ. Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat Genet* 1999; 22: 94–97.
94. Amedeo P, Habu Y, Afsar K, Scheid OM, Paszkowski J. Disruption of the plant gene MOM releases transcriptional silencing of methylated genes. *Nature* 2000; 405: 203–206.

Copyright © 2001 John Wiley & Sons, Ltd.

TAB 15

This material may be protected by Copyright law (Title 17 U.S. Code)

Cell, Vol. 32, 1105–1113, April 1983, Copyright © 1983 by MIT

0092-8674/83/041105-09 $02.00/0

Independent Mechanisms Involved in Suppression of the Moloney Leukemia Virus Genome during Differentiation of Murine Teratocarcinoma Cells

Ohtsura Niwa* Yoshifumi Yokota,*
Hiroyuki Ishida* and Tsutomu Sugahara†
* Department of Experimental Radiology
Faculty of Medicine
Kyoto University
Sakyo-ku, Kyoto 606, Japan
† Kyoto National Hospital
Fushimi-ku, Kyoto 612, Japan

Summary

Expression and DNA methylation of the Moloney murine leukemia virus (M-MuLV) genome were investigated in murine teratocarcinoma cells after virus infection. The newly acquired viral genome was devoid of methylation, yet its expression was repressed. The integrated viral genome in undifferentiated teratocarcinoma cells was methylated within 15 days after infection. Although 5-azacytidine decreased the level of DNA methylation, it did not activate M-MuLV in undifferentiated cells. Activation by 5-azacytidine occurred only in differentiated teratocarcinoma cells. Thus two independent mechanisms seem to regulate gene expression during the course of differentiation. The first mechanism operates in undifferentiated cells to block expression of M-MuLV and other exogeneously acquired viral genes, such as SV40 and polyoma virus, and does not depend on DNA methylation. The second mechanism relates only to differentiated cells and represses expression of genes in which DNA is methylated.

Introduction

Laboratory strains of mice as well as field isolates of Mus species carry a large number of endogenous type C virus genomes. Blot hybridization analysis of mouse DNA digested with various restriction enzymes suggested that cells of inbred mice have as many as 50 copies of DNA sequences related to type C viruses (Dolberg et al., 1981). Inducible murine endogenous viruses are classified into two host range groups, ecotropic viruses and xenotropic viruses. Chromosomal locations of ecotropic viruses can be determined by classical Mendelian genetics. Ecotropic virus loci of AKR mice, *Akv1* and *Akv2*, have been mapped on chromosome 7 (Rowe et al., 1972) and on chromosome 16 (Kozak and Rowe, 1980), respectively. The *Cv* locus, an ecotropic virus locus of C3H/HeJ and BALB/c strains was detected on chromosome 5 (Kozak and Rowe, 1979; Ihle et al., 1979). C3H/Fg mice have an ecotropic virus locus on chromosome 7 at a site distinct from the *Akv1* locus. C58 and B10·BR mice carry unnamed ecotropic virus loci on chromosome 8 and on chromosome 11, respectively. Thus the distribution of endogenous ecotropic virus loci over a wide range of mouse chromosomes suggests that these endogenous virus genomes may have recently derived from a common prototype virus through germline integration during establishment of laboratory mice. Actually, Jaenisch demonstrated that infection of preimplantation stage BALB/129 mouse embryos with M-MuLV leads to integration of the virus into the germ line, in an endogenous unexpressed state (Jaenishch, 1976).

Infection of mouse embryo fibroblasts with murine ecotropic viruses usually results in stable integration of the viral genome and subsequent production of progeny virus. This makes for a clear distinction between fibroblast lines and the cells of preimplantation stage embryos where infection with exogenous viruses results in repression of the integrated virus genomes. Murine teratocarcinoma cells are the malignant and pluripotent stem cells derived from carcinoma of preimplantation stage embryos (Lehman et al., 1974). Teratocarcinoma stem cells infected with ecotropic murine type C viruses do not produce progeny virus (Peries et al., 1977; Teich et al., 1977; Huebner et al., 1979; Gautsch, 1980). The infecting viral genome, however, is carried in these cells and can be activated with bromodeoxyuridine (BrdUrd) (Speers et al., 1980).

The expression of endogenous virus genomes in avian and mouse embryo fibroblasts is regulated by DNA modification and can be activated by treating the cells with 5-azaCytidine (5-AzaCyd), a potent inhibitor of DNA methylation (Groudine et al., 1981; Niwa and Sugahara, 1981). We studied the state of the M-MuLV genome in teratocarcinoma cells and found that the viral genome in undifferentiated stem cells is repressed not by DNA methylation but by some other mechanism. Only after differentiation of the cells is the viral sequence under the control of a mechanism regulated by DNA modification.

Results

Repression of M-MuLV in Undifferentiated Teratocarcinoma Cells, and Induction of Its Expression by BrdUrd and Retinoic Acid

EC-A1 cells derived from PCC4 cells, a pluripotent stem cell line of mouse teratocarcinoma cells, were infected with M-MuLV. Cells were treated for 24 hr with BrdUrd and/or retinoic acid, a compound that induces differentiation of teratocarcinoma cells (Strickland and Mahdavi; 1978) before or after virus infection. Since both BrdUrd and retinoic acid exerted potent cytotoxic effects on EC-A1 cells, the virus-producing cells in the drug-treated cultures may be preferentially eliminated to yield a false negative result. Thus EC-A1 cells were treated with BrdUrd and/or retinoic acid before or after virus infection, and the cells were then cocultivated with SC-1 cells. The SC-144/264cells were passaged once and assayed for virus

expression. The results in Table 1 indicate that preinfection and postinfection treatment with BrdUrd alone was effective for productive infection of the virus. Preinfection treatment with retinoic acid made the EC-A1 cells susceptible to infection with M-MuLV. The expression of the virus in the BrdUrd-treated cells was enhanced by the presence of retinoic acid. Although less efficient, postinfection treatment of the cells with BrdUrd together with retinoic acid did induce expression of the virus. A low but definite production of M-MuLV was detected by amplification through SC-1 cells when the EC-A1 cells were treated for 24 hr with 0.3–1 M of retinoic acid alone, immediately after infection.

Infection of F-9 cells, a nullipotent teratocarcinoma cell line (Bernstine et al., 1973), with M-MuLV also required preinfection or postinfection treatment with BrdUrd for efficient expression of the virus (O. Niwa, unpublished observation).

The teratocarcinoma stem cells, EC-A1 and F-9 cells, were all derived from 129 mice that contained the ecotropic virus sequences and were mostly subgenomic in size (Chan et al., 1980). These have been classified as a no virus strain (Chattopadhyay et al., 1974). The virus which was recovered from BrdUrd-treated EC-A1 cells, and which had been infected with M-MuLV, grew equally well on NIH/3T3 and BALB/3T3 cells. The tropism of the virus and the lack of virus activation from uninfected EC-A1 cells (O. Niwa, unpublished observation) strongly suggest that the virus thus recovered from infected EC-A1 cells after BrdUrd treatment is M-MuLV and not the virus endogenous to EC-A1 cells.

Table 1. Plaque-Forming Cells per 2×10^5 SC-1 Cells Cocultivated with EC-A1 Cells Treated Preinfection and Postinfection with BrdUrd and/or Retinoic Acid

BrdUrd (μg/ml)	Retinoic Acid (μM)	Preinfection Treatment/ Postinfection Treatment
0	0	0/0
0	0.03	5/0
0	0.1	14/0
0	0.3	23/5
0	1.0	46/3
5	0	122/7
5	0.03	634/7
5	0.1	TMTC/25
5	0.3	TMTC/31
5	1.0	TMTC/23
10	0	TMTC/6
10	0.03	TMTC/15
10	0.1	TMTC/43
10	0.3	TMTC/22
10	1.0	TMTC/19
20	0	TMTC/12
20	0.03	TMTC/21
20	0.1	TMTC/48
20	0.3	TMTC/39
20	1.0	TMTC/51

TMTC: too numerous to count.

Isolation of EC-A1(Mo) Clones Carrying M-MuLV Genome in a Repressed State

EC-A1 cells were infected with M-MuLV at a moi of 1.0, which had been determined by titration on SC-1 cells. The cultures were then trypsinized, and the cells were plated for ring cloning. Each of the randomly isolated clones was treated with 20 μg/ml of BrdUrd and cocultivated with SC-1 cells. SC-1 cells were passaged three times and tested for virus expression by the reverse XC test. Of 110 clones thus tested, 47 expressed virus after BrdUrd treatment. These virus-inducible clones, designated EC-A1(Mo) clones, therefore carry the M-MuLV genome, in a repressed form.

Transcriptional Control of M-MuLV Expression in Undifferentiated Teratocarcinoma Cells

Liquid hybridization experiments indicated no detectable viral RNA transcript in M-MuLV-infected teratocarcinoma cells (Teich et al., 1977). We have isolated total cellular RNA from one of the virus-carrying clones, EC-A1(Mo)4. RNA was size-differentiated on agarose gel electrophoresis, transferred to diazobenzyloxymethyl paper (DBM paper) and analyzed for the sequence hybridizable to ^{32}P-labeled M-MuLV DNA (Figure 1). SC-1 cells productively infected with M-MuLV had two major bands, 34S and 24S, corresponding to virus genomic RNA and spliced *env* gene messenger RNA. On the other hand, uninfected SC-1 cells, EC-A1 cells and EC-A1(Mo)4 cells contained no detectable level of RNA sequence hybridizable to M-MuLV DNA. The same filter was hybridized with an 18S rRNA probe. Three bands corresponding to 18S rRNA and two precursor RNAs were detected in all four RNA samples. The density of the bands revealed by this probe varied little among the four cell lines tested, indicating that the amounts of RNA used for the analysis were the same for these cell lines. Total cellular RNA isolated from EC-A1 cells 48 hr after infection with M-MuLV was analyzed by the dot blot hybridization technique (data not shown). Here too, the RNA transcript of M-MuLV was not detected.

5-AzaCyd Induction of M-MuLV Gene Expression and the State of Differentiation of Teratocarcinoma Cells

We attempted to induce M-MuLV expression in undifferentiated EC-A1(Mo) clones by 5-AzaCyd. To our surprise, none of the 47 EC-A1(Mo) clones expressed virus after treatment with 2 μg/ml 5-AzaCyd, although virus was readily recovered when the clones were treated with BrdUrd.

To determine whether or not the DNA methylation was suppressed by 5-AzaCyd in undifferentiated teratocarcinoma cells, two clones, EC-A1(Mo)1 and EC-A1(Mo)4, were treated with 5-AzaCyd, and the level of DNA methylation was measured. As is clear from

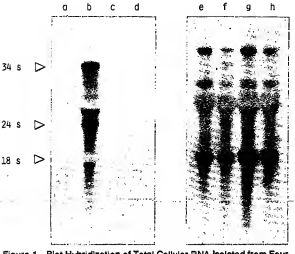


Figure 1. Blot Hybridization of Total Cellular RNA Isolated from Four Cell Lines

Probes used were ^{32}P-labeled M-MuLV DNA (lanes a, b, c and d) and ^{32}P-labeled cloned rRNA gene (lanes e, f, g and h). (Lanes a and e) RNA from SC-1 cells; (lanes b and f) RNA from SC-1 cells infected with M-MuLV; (lanes c and g) RNA from EC-A1 cells; (lanes d and h) RNA from EC-A1(Mo)4 cells.

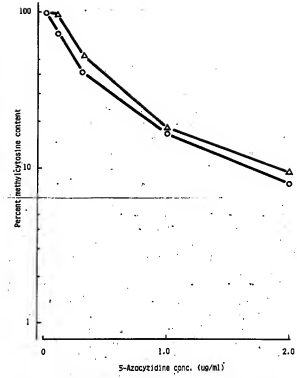


Figure 2. Level of Methylcytosine in Newly Replicated DNA of EC-A1 Cells

Cells were seeded onto 3 cm dishes at a concentration of 1 × 10^6 cells per dish. After overnight incubation, the dishes were nourished with medium containing (methyl-^{3}H)methionine at 4 μCi/ml and (2-^{14}C)thymidine at 0.005 μCi/ml. DNA was collected, hydrolyzed and separated on a cellulose thin-layer glass plate. The ratio of ^{3}H and ^{14}C counts in methylcytosine and thymidine respectively was taken as a relative measure of methylcytosine content (○) EC-A1(Mo)1 cells; (△) EC-A1(Mo)4 cells.

Figure 2, the level of methylcytosine in cellular DNA decreased with increasing doses of 5-AzaCyd. Therefore, in undifferentiated teratocarcinoma cells, suppression of DNA methylation does not seem to result in activation of the viral genome.

EC-A1(Mo)4 cells were treated with dimethylacetamide for 2 weeks to induce differentiation, and such was assessed to be complete by the morphology. M-MuLV was never expressed in the differentiated EC-A1(Mo)4 cells. These cells were then treated with 5-AzaCyd and cocultivated with SC-1 cells. The SC-1 cells were passaged twice, and the reverse XC test was performed at each passage. Only in the differentiated EC-A1(Mo)4 cells was the 5-AzaCyd-activated expression of M-MuLV evident (Table 2). Therefore, at least in the differentiated EC-A1(Mo)4 cells, the viral genome seems to be regulated by the extent of DNA methylation, and suppression of DNA methylation activates expression of the virus.

Table 2. Plaque-Forming Cells per 2 × 10^6 SC-1 Cells Cocultivated with EC-A1(Mo)4 Cells Treated with 5-AzaCyd

Cells	Concentration of 5-AzaCyd (μg/ml)	First Passage/ Second Passage
EC-A1(Mo)4 cells undifferentiated	2	0/0
	4	0/0
	6	0/0
	8	0/0
EC-A1(Mo)4 cells differentiated	1	0/0
	2	2/31
	4	13/TMTC

TMTC: too numerous to count.

Lack of DNA Methylation of Unintegrated M-MuLV Genome

The data presented above indicate that the M-MuLV genome in the differentiated teratocarcinoma cells is methylated. Experiments were then designed to determine the timing of DNA methylation of the M-MuLV genome, after virus infection of the undifferentiated cells.

EC-A1 cells were infected with M-MuLV, and DNA was extracted from the Hirt supernatant fraction which contained unintegrated viral genome. DNA isolated from EC-A1 cells 6 hr after infection contained three molecular species of M-MuLV—namely, closed circular, linear and open circular DNAs (Figure 3, lane a). Closed circular DNA and open circular DNA consisted of two subbands differing slightly in size. A linear molecule had a single band of 8.8 kb. Upon cleavage with Hind III, two bands corresponding to 8.2 kb and 8.8 kb linear molecules were detected (Figure 3, lane b) and were assumed to be the full-sized M-MuLV genome with one and two long terminal repeats (LTRs). All of these sequences were com-

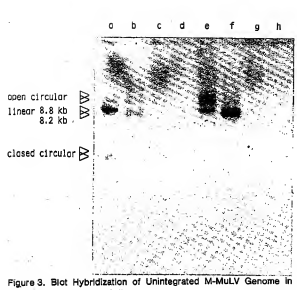


Figure 3. Blot Hybridization of Unintegrated M-MuLV Genome in Undifferentiated EC-A1 Cells

EC-A1 cells were infected with M-MuLV. DNA was extracted from the Hirt supernatant fraction of EC-A1 cells 8 hr after infection with M-MuLV (lanes a, b, c and d), or 12 hr after infection (lanes e, f, g and h). The probe used was ^{32}P-labeled M-MuLV DNA. (Lanes a and e) undigested DNA; (lanes b and f) digestion with Hind III; (lanes c and g) digestion with Hap II; (lanes d and f) digestion with Msp I.

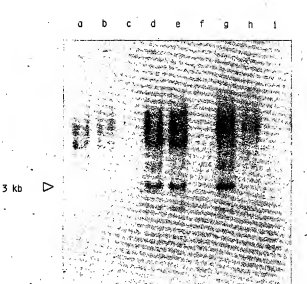


Figure 4. Blot Hybridization of Newly Integrated M-MuLV DNA in Undifferentiated EC-A1 Cells

DNA was isolated from uninfected EC-A1 cells (lane a, b and c), EC-A1(Mo)4 cells (lane d, e and f) or EC-A1 cells 48 hr after infection with M-MuLV (lanes g, h and i). The probe used was the ^{32}P-labeled Sma I fragment of M-MuLV DNA. (Lanes a, d and g) digestion with Bam HI; (lanes b, e and h) digestion with Bam HI plus Hap II; (lanes c, f and i) digestion with Bam HI plus Msp I.

pletely digested by Hpa II and Msp I, suggesting that they are devoid of methylation at the CCGG sequence (Figure 3, lanes c, d). Similar results were obtained with DNA isolated at 12 hr after infection (Figure 3, lanes e, f, g, h). However, one difference in the 12 hr sample was that this DNA lacked a closed circular molecule (Figure 3, lane e). It is of interest that the Hirt supernatant DNA isolated 24 hr after infection had a much lesser amount of viral DNA, while the 48 hr sample was devoid of M-MuLV sequences (data not shown). In the case of SC-1 cells, DNA extracted 24 hr after infection had the greatest amount of unintegrated M-MuLV provirus sequence (O. Niwa, unpublished observation).

Undermethylation of Freshly Integrated M-MuLV Genome and Its Subsequent Methylation during Multiple Cell Cycling of Undifferentiated Cells

EC-Al cells were infected with M-MuLV at a moi of 2, and DNA was isolated 48 hr later. The integration of viral genome occurs at random sites. Digestion of DNA from randomly infected cells with Eco RI, which does not cleave the M-MuLV genome, will produce multiple fragments of various sizes carrying integrated viral sequences flanked by cellular sequences, and these cannot be resolved by agarose gel electrophoresis. Therefore, we used Bam HI, which cut the M-MuLV genome internally. As a probe to detect viral genome, the 316 base Sma I fragment of M-MuIV DNA was labeled with ^{32}P. This probe allows for detection of the 3 kb internal fragment of Bam HI–digested M-MuLV together with other sequences derived from endogenous viruses (Figure 4, lanes a, d, g). Uninfected EC-A1 cells lacked the 3 kb M-MuLV-specific sequence (Figure 4, lane a). DNA of EC-A1 cells 48 hr after infection contained the 3 kb band (Figure 4, lane g), and this fragment was not derived from the unintegrated M-MuLV, as unintegrated provirus DNA disappeared from the Hirt supernatant fraction of EC-A1 cells before 48 hr of infection. Double digestion of 48 hr postinfection DNA with Hap II and with Msp I completely eliminated the 3 kb Bam HI band, suggesting that this part of the M-MuLV genome is devoid of DNA methylation (Figure 4, lanes h, i). DNA was isolated from EC-A1(Mo)4 cells that had undergone multiple cell cycling after virus infection. The M-MuLV-specific 3 kb Bam HI band was clearly demonstrable in the cellular DNA (Figure 4, lane d). However, the same sequence was now resistant to Hap II digestion (Figure 4, lane e). Other bands derived from endogenous virus genomes were also resistant to Hap II (Figure 4, lanes b, e, h). DNAs from another EC-A1(Mo) clone, EC-A1(Mo)22, and randomly infected EC-A1 cells passaged at least 20 times also contained the Hpa II–resistant 3 kb Bam HI fragment specific to M-MuLV.

Experiments were performed to determine the time of the DNA methylation of M-MuLV genome after infection. DNA was isolated from M-MuLV-infected EC-A1 cells on days 4, 10 and 15 after infection, and analyzed for the state of DNA methylation (Figure 5). The M-MuLV genome was devoid of methylation at

Hap II sites up to day 10 (Figure 5, lanes b, e). However, the viral genome was resistant to Hap II when it was isolated from 15-day-old cultures. Therefore, the integrated M-MuLV genome was methylated between days 10 and 15 in culture. Production of M-MuLV from infected EC-A1 cells during these 10 days was nil.

These results indicate that although unintegrated and newly integrated M-MuLV genomes were devoid of DNA methylation, the same sequence is methylated in the undifferentiated cells kept in culture for over 15 days.

Lack of Change in the State of DNA Methylation of the M-MuLV Genome during Differentiation of Teratocarcinoma Cells

DNA was isolated from undifferentiated EC-A1(Mo)4 cells and dimethylacetamide-induced differentiated EC-A1(Mo)4 cells. These DNAs contained the 3 kb Bam HI fragment of the integrated M-MuLV genome, as well as other bands derived from endogenous virus genomes (Figure 6). The 3 kb fragment of M-MuLV in undifferentiated EC-A1(Mo)4 cells was again resistant to digestion with Hap II, confirming the result in Figure 4, lane e (Figure 6, lane b). Similar resistance to Hap II enzyme was noted when the DNA from differentiated EC-A1(Mo)4 cells was analyzed. Therefore, the pattern of DNA methylation of at least the 3 kb Bam HI fragment did not change during differentiation of the cells, yet inducibility of the viral genome by 5-AzaCyd treatment changed drastically.

Transfection with DNA from M-MuLV Infected Teratocarcinoma Cells

DNA could be isolated from EC-A1 cells 2 days after infection with M-MuLV and also from EC-A1(Mo) clones. SC-1 cells were transfected with these DNAs (Table 3). Although at a low efficiency, DNA from EC-A1 cells infected with M-MuLV 2 days previously was capable of producing M-MuLV, while DNA from EC-A1(Mo) clones was not.

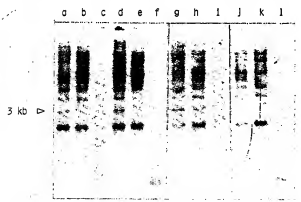


Figure 5. Blot Hybridization of Newly Integrated M-MuLV DNA in Undifferentiated EC-A1 Cells

DNA was isolated from EC-A1 cells on the day 4 (lanes a, b and c), day 10 (lanes d, e and f) or day 15 (lanes g, h and i) after infection with M-MuLV, or from uninfected EC-A1 cells (lanes j, k and l). (Lanes a, d, g and j) digestion with Bam HI; (lanes b, e, h and k) digestion with Bam HI plus Hap II; (lanes c, f, i and l) digestion with Bam HI plus Msp I.

Discussion

Jaenisch and coworkers have demonstrated that endogenous virus can be formed by infection of the preimplantation stage embryo cells, with exogenous virus (Jaenisch et al., 1975; Jaenisch, 1976). Infection of undifferentiated stem cells of mouse teratocarcinoma leads to silencing of the exogenously acquired

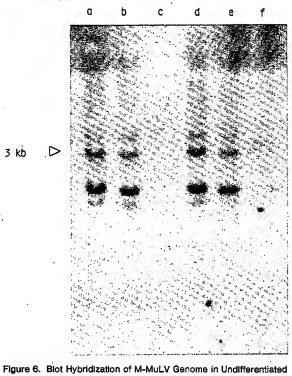


Figure 6. Blot Hybridization of M-MuLV Genome in Undifferentiated and Differentiated EC-A1(Mo)4 Cells

DNA was isolated from undifferentiated (lanes a, b and c) and differentiated EC-A1(Mo)4 cells (lanes d, e and f). The probe used was a ^{32}P-labeled 1.65 kb Bam HI–Hind III fragment of M-MuLV DNA. (Lanes a and d) digestion with Bam HI; (lanes b and e) digestion with Bam HI plus Hap II; (lanes c and f) digestion with Bam HI plus Msp I.

Table 3. Transfection of SC-1 Cells with DNA from M-MuLV-Infected Cells

DNA Source	DNA per Dish (μg)	Virus-Positive/ Total
SC-1	20	0/20
SC-1 M-MuLV-infected	20	23/30
EC-A1	20	0/20
EC-A1(Mo)4	20	0/20
EC-A1(Mo)22	20	0/20
EC-A1(Mo)26	20	0/20
EC-A1 M-MuLV-infected[a]	20	9/20

[a] DNA was isolated 2 days after infection with M-MuLV.

type C virus genome (Teich et al., 1977; Speers et al., 1980). However, the mechanism of repression of integrated viral genome in the cells of the preimplantation stage embryo and in teratocarcinoma cells has remained unknown.

Undifferentiated teratocarcinoma cells have characteristics of preimplantation stage embryo cells in that they lack H-2 antigen on the cell surface (Artzt and Jacob, 1974), possess two active X chromosomes in female cells (Martin et al., 1978; McBurney and Strutt, 1980) and have the potential to differentiate into a variety of cell types (Kleinsmith and Pierce, 1964; Mintz and Illmensee, 1975). Thus teratocarcinoma cells are a pertinent model for studies on the undifferentiated state of embryogenesis.

For cells to be in an undifferentiated state, they must be equipped with the potential to suppress specifically expression of genes required only for differentiated cells. Mouse teratocarcinoma cells are resistant to exogenously incorporated genetic elements such as SV40 and polyoma virus (Swartzendruber and Lehman, 1975; Swartzendruber et al., 1977; Segal and Khoury, 1979) as well as to murine leukemia virus. These viral genomes may be regarded as unnecessary luxury genes in undifferentiated teratocarcinoma cells.

We found that the unintegrated M-MuLV genome and at least part of the newly integrated genome, the 3 kb Bam HI fragment, in undifferentiated teratocarcinoma cells are devoid of DNA methylation, yet the expression of the genome is repressed. SV40 DNA was also shown to be undermethylated in teratocarcinoma cells (Friedrich and Lehman, 1981). Treatment with retinoic acid, which triggers differentiation of the cells, can activate virus expression in undifferentiated cells, provided that drug treatment follows shortly after infection with M-MuLV, when methylation of the M-MuLV genome has not yet occurred (Table 1). However, the frequency of virus expression induced by retinoic acid was rather low, and for the entire course of differentiation about 10 days were required. The low frequency thus observed might be due to the incompleteness of differentiation. The finding that the freshly integrated M-MuLV genome in undifferentiated teratocarcinoma cells is transfectionally active on SC-1 cells (Table 3) suggests that this genome could be transcriptionally active in differentiated cells. Therefore, the M-MuLV genome freshly acquired by EC-A1 cells seems to be devoid of DNA modification, and DNA modification is known to reduce the rate of transcription (Stuhlmann et al., 1981; Hoffmann et al., 1982). Actually, the M-MuLV genome was devoid of methylation at Hpa II sites for at least 10 days after infection; it was not expressed in the undifferentiated cells. The 3 kb fragment of M-MuLV genome is subsequently methylated in infected cells kept in culture for over 15 days. The virus genome becomes transfectionally inactive and cannot be induced by simple differentiation of the cells. Also, suppression of DNA methylation by 5-AzaCyd did not activate expression of the virus genome in EC-A1(Mo) clones. Dot blot hybridization analysis of RNA isolated from 5-AzaCyd-treated EC-A1(Mo)4 cells that carry the methylated M-MuLV genome did not contain RNA hybridizable to M-MuLV probe, while in the BrdUrd-treated culture there was a marked increase of M-MuLV transcript (O. Niwa, unpublished observation). Therefore, we conclude that the M-MuLV genome is repressed in the undifferentiated cells by a mechanism other than DNA methylation. This repression can be unblocked by the treatment of the cells with BrdUrd. EC-A1 cells freshly infected with M-MuLV, and carrying the unmethylated viral genome, and EC-A1(Mo) clones presumably carrying the methylated M-MuLV genome were both induced by BrdUrd treatment (see Table 1 and above). Incorporation of BrdUrd into DNA does not affect the level of methylation (unpublished observations).

DNA methylation has no effect on the transcriptional activity of the M-MuLV genome in undifferentiated cells. The M-MuLV genome is nevertheless methylated in cells that have undergone multiple cell cycling. Genes that are not transcribed may be preferentially methylated in undifferentiated cells.

The pattern of DNA methylation of the 3 kb Bam HI fragment did not change during differentiation of EC-A1(Mo)4 cells. However, the M-MuLV genome is now susceptible to induction by 5-AzaCyd. Repression of the viral genome by DNA methylation thus seems operative only in differentiated cells. DNA methylation is thought to suppress gene expression through a condensation of chromatin, and antisera raised against methylcytosine bind to the heterochromatic region of the mouse chromosome (Miller et al., 1974). Methylcytosine is more abundant in the fraction of chromation that is resistant to DNAase digestion (Razin and Ceder, 1977). Methylated sequences of endogenous viruses are located on chromatin regions that are resistant to DNAase I (van der Putten et al., 1982).

Cells of the preimplantation stage embryo have two active X chromosomes and lack heterochromatin (Epstein et al., 1978). Heterochromatinization of an X chromosome occurs during differentiation of teratocarcinoma cells (Martin et al., 1978; McBurney and Strutt, 1980). Condensation of the heterochromatic region may even be facilitated by the presence of yet unidentified chromatin protein(s), and differentiation of cells may trigger production of this chromatin protein, which inactivates already methylated luxury genes by condensation of their chromatin regions.

In light of all these data, we propose the presence of two independent mechanisms regulating gene expression in mammalian cells. The first mechanism which is operating in undifferentiated cells, is not influenced by the state of DNA methylation for its function. Repression of transcription by this mecha-

nism may involve discrimination against the luxury gene promoter. Host range mutants of polyoma virus that can replicate on undifferentiated teratocarcinoma cells were found to possess mutations at the promoter region (Sekikawa and Levine, 1981). The second mechanism suppresses the expression of methylated genes by changing the conformation of their chromatin domains, and this mechanism operates only in differentiated cells. When M-MuLV infects undifferentiated teratocarcinoma cells, the viral genome is suppressed by the first mechanism. The transcriptionally inactive genome of M-MuLV may be gradually methylated during replication of the host cells. Differentiation of the cells terminates regulation by the first mechanism and activates the second mechanism, which now recognizes the methylated genome of M-MuLV and represses its expression by condensation of the chromatin domain. 5-AzaCyd unblocks the second mechanism by decreasing the level of methylcytosine, while BrdUrd unblocks both mechanisms. Since BrdUrd is known to change the affinity of DNA-binding proteins for DNA (Lin and Riggs, 1972, 1976; Goeddel et al., 1977), binding of some chromatin protein(s) to the promoter region of M-MuLV may be responsible for operation of the first mechanism. The first mechanism of the repression of gene expression seems to be *trans*-acting, since the extrachromosomal genome of SV40 and the M-MuLV genome integrated at random sites in undifferentiated cells are repressed. These findings suggest that the first mechanism may inactivate gene expression, possibly by some diffusible repressor-like protein, and if such is the case, the *trans*-acting nature can be readily explained.

During preparation of this manuscript, a report appeared on a similar subject (Stewart et al., 1982) These authors' data suggest that the M-MuLV genome becomes methylated immediately after integration. This difference between our results and theirs might relate to different cell lines. We used PCC4-derived EC-A1 cells, while they used F-9 cells. Gautsch and Wilson (1983) apparently obtained findings similar to those we report here.

Experimental Procedures

Cell Lines and Viruses

EC-A cells (Gautsch, 1980), a subline of PCC4 Aza1 cells (Jakob et al., 1973), were kindly provided by J. Gautsch. EC-A1 cells, a subclone of EC-A cells isolated in our laboratory, were also used. F-9 cells (Strickland and Mahdavi, 1978) were obtained from K. Sekigawa. SC-1 cells, a mouse embryo fibroblast line derived from a feral mouse (Hartley and Rowe, 1975), were obtained from A. Declève. Teratocarcinoma cell lines were grown in Eagle's minimal essential medium supplemented with 10% heat-inactivated fetal calf serum. SC-1 cells were grown in MEM alpha medium (Gibco) supplemented with 10% heat-inactivated calf serum.

M-MuLV was obtained from A. Ishimoto and grown on SC-1 cells.

Virus Infection, Transfection and Biological Assays

Procedures of virus infection were as described previously (Niwa et al., 1973). Titration of M-MuLV stock was on SC-1 cells by UV-XC assay (Rowe et al., 1970). M-MuLV-infected teratocarcinoma cells were assayed directly or after cocultivation with SC-1 cells by the reverse XC cell assay (Niwa et al., 1973).

Transfection of SC-1 cells with DNA from M-MuLV-infected EC-A1 cells was carried out by a procedure described by other workers (Stuhlmann et al., 1981).

Drug Treatments

Stock solutions of retinoid and BrdUrd were prepared in dimethylsulfoxide at concentrations of 10 mM and 1 mg/ml, respectively, and kept in the dark at −20°C. Teratocarcinoma cells were incubated with these drugs for 24 hr at 37°C either before or after infection with M-MuLV. When cocultivation with SC-1 cells was used for the amplification of progeny virus, teratocarcinoma cells were treated for 30 min with 25 μg/ml mitomycin C to suppress overgrowth of the undifferentiated cells.

Differentiated teratocarcinoma cells were obtained by the procedure of Speers et al. (1980) with slight modification. Undifferentiated teratocarcinoma cells grown as monolayer cultures were nourished every 3 days with a medium containing 10 mM dimethylacetamide for 2 weeks. Cultures consisted only of cells with an epithelial morphology. Although retinoic acid also induced differentiation of EC-A1 cells, it was more cytotoxic than dimethylacetamide.

Base Analysis

The level of methylcytosine in the newly synthesized DNA was assayed as described previously (Niwa and Sugahara, 1981).

Extraction of Cellular RNA and DNA

Total cellular RNA was isolated by sedimentation through cesium chloride as described by others (Chirgwin et al., 1979).

For the isolation of DNA, dishes were washed with phosphate-buffered saline solution and digested at 37°C for 2 hr with 100 μg/ml RNAase A in 1% sodium dodecylsulfate, 0.1 M NaCl, 5 mM EDTA and 20 mM Tris–HCl (pH 8.0). Proteinase K was then added at 100 μg/ml, and dishes were incubated for another 2 hr at 37°C. The lysate was extracted three times with phenol–chloroform and ethanol-precipitated. DNA thus extracted was used for further analysis. Unintegrated viral DNA was isolated by the procedure of Hirt (1967).

Restriction Endonuclease Digestion, Gel Electrophoresis and Hybridization

Restriction enzymes were obtained from Takara Shuzo Co., Ltd. (Kyoto, Japan) except for Msp I, which was purchased from New England Biolaboratory. After cleavage with restriction enzymes, DNA samples were electrophoresed on 0.7% horizontal agarose gels and transferred to a nitrocellulose sheet, as described by Southern (1975).

Total cellular RNA was denatured, as described (McMaster and Carmichael, 1977), electrophoresed on 1.1% agarose gels and transferred to DBM paper accordingly (Alwine et al., 1977). For the dot blot hybridization, nondenatured RNA was spotted directly onto nitrocellulose filter (Thomas, 1980).

The recombinant plasmid which carries the 8.2 kb M-MuLV genome at the Hind III site and which was cloned by J. W. Hoffmann, was a generous gift from R. A. Weinberg. The DNA fragment of 8.2 kb containing the entire sequence of M-MuLV was purified by agarose gel electrophoresis. M-MuLV DNA was further digested by Bam HI and Sma I. The Hind III–Bam HI fragment of 1.65 kb was recovered from agarose gel after electrophoresis. The Sma I fragment of 316 bases was electrophoresed through a polyacrylamide gel. Recombinant plasmids carrying 18S and 28S rRNA gene of the mouse were a kind gift from R. Kominami (Urano et al., 1980; Kominami et al., 1982). These DNAs were labeled by the nick translation procedure (Maniatis et al., 1975). Specific activities of the probes were 1–4 × 10^8 cpm/μg.

Acknowledgments

We thank Drs. R. A. Weinberg and J. W. Hoffmann for the generous gift of recombinant plasmid, M. Yoshida for pertinent advice on biochemical techniques, A. Ishimoto for providing M-MuLV stock,

Prof. S. Numa, in whose laboratory part of this investigation was carried out, R. Kominami for provision of plasmids carrying rRNA genes, K. Sekikawa from whom we obtained the F-9 cells, M. Ohara for reading the manuscript and Y. Fukuda for secretarial services. This work was supported by a grant-in-aid from the Ministry of Education, Science, and Culture, Japan.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 14, 1982; revised January 25, 1983

References

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Nat. Acad. Sci. USA *74*, 5350–5354.

Artzt, K. and Jacob, F. (1974). Absence of serologically detectable H-2 on primitive teratocarcinoma cells in culture. Transplantation *17*, 632–634.

Bernstine, E. G., Hooper, M. L., Granchamp, S. and Ephrussi, B. (1973). Alkaline phosphatase activity in mouse teratoma. Proc. Nat. Acad. Sci. USA *70*, 3899–3903.

Chan, H. W., Bryan, T., Moore, J. L., Staal, S. P., Rowe, W. P. and Martin, M. A. (1980). Identification of ecotropic proviral sequences in inbred mouse strains with a cloned subgenomic DNA fragment. Proc. Nat. Acad. Sci. USA *77*, 5779–5783.

Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. and Rowe, W. P. (1974). Qualitative and quantitative studies on AKR-type murine leukemia virus sequences in mouse DNA. Cold Spring Harbor Symp. Quant. Biol. *39*, 1085–1101.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Dolberg, D. S., Bacheler, L. T. and Fan, H. (1981). Endogenous type C retroviral sequences of mice are organized in a small number of virus-like classes and have been acquired recently. J. Virol. *40*, 96–106.

Epstein, C. J., Smith, S., Travis, B. and Tucker, G. (1978). Both X chromosomes function before visible X-chromosome inactivation in female mouse embryos. Nature *274*, 500–503.

Friedrich, T. D. and Lehman, J. M. (1981). The state of similar virus 40 DNA in the embryonal carcinoma cells of the murine teratocarcinoma. Virology *110*, 159–166.

Gautsch, J. W. (1980). Embryonal carcinoma stem cells lack a function required for virus replication. Nature *285*, 110–112.

Gautsch, J. W. and Wilson, M. (1983). Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature, in press.

Goeddel, D. V., Yansura, D. G. and Caruthers, M. H. (1977). Binding of synthetic lactose operator DNAs to lactose repressors. Proc. Nat. Acad. Sci. USA *74*, 3292–3296.

Groudine, M., Eisenman, R. and Weintraub, H. (1981). Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature *292*, 311–317.

Hartley, J. W. and Rowe, W. P. (1975). Clonal cell line from a feral mouse embryo which lacks host range restriction for murine leukemia virus. Virology *65*, 128–134.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. *26*, 365–369.

Hoffman, J. W., Steffen, D., Gusella, J., Tabin, C., Bird, S., Cowing, D. and Weinberg, R. A. (1982). DNA methylation affecting the expression of murine leukemia proviruses. J. Virol. *44*, 144–157.

Heubner, K., Tsuchida, N., Green, C. and Croce, C. M. (1979). A murine teratocarcinoma stem cell line carries suppressed oncogenic virus genomes. J. Exp. Med. *150*, 392–405.

Ihle, J. N., Joseph, D. R. and Domotor, J. J. (1979). Genetic linkage of C3H/HeJ and BALB/c endogenous ecotropic C-type viruses to phosphoglucomutase-1 on chromosome 5. Science *204*, 71–73.

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc. Nat. Acad. Sci. USA *73*, 1260–1264.

Jaenisch, R., Fan, H. and Croker, B. (1975). Infection of preimplantation mouse embryos and of new born mice with leukemia virus: tissue distribution of viral DNA and RNA leukemogenesis in the adult animal. Proc. Natl. Acad. Sci. USA *72*, 4008–4012.

Jakob, H., Boon, T., Gillard, J., Nicholas, J. F. and Jacob, F. (1973). Teratocarcinoma da la souris: isolement, culture et proprietes de cellules a potentialites multiples. Ann. Microbiol. (Paris). *124B*, 269–282.

Kleinsmith, L. J. and Pierce, G. B., Jr. (1964). Multipotentiality of single embryonal carcinoma cells. Cancer Res. *24*, 1544–1551.

Kominami, R., Mishima, Y., Urano, Y., Sakai, M. and Muramatsu, S. (1982). Cloning and determination of the transcription termination site of ribosomal RNA gene of the mouse. Nucl. Acids Res. *10*, 1963–1979.

Kozak, C. A. and Rowe, W. P. (1979). Genetic mapping of the ecotropic murine leukemia virus-inducing locus of BALB/c mouse to chromosome 5. Science *204*, 69–71.

Kozak, C. A. and Rowe, W. P. (1980). Genetic mapping of the ecotropic virus-inducing locus *Akv-2* of the AKR mouse. J. Exp. Med. *152*, 1419–1423.

Lehman, J. M., Speers, W. C., Swartzandruber, D. E. and Pierce, G. B. (1974). Neoplastic differentiation: characteristics of cell lines derived from a murine teratocarcinoma. J. Cell. Physiol. *84*, 13–28.

Lin, S. Y. and Riggs, A. (1972). *Lac* operator analogues: bromodeoxyuridine substitution in the *lac* operator affects the rate of dissociation of the *lac* repressor. Proc. Nat. Acad. Sci. USA *69*, 2574–2576.

Lin, S. Y. and Riggs, A. (1976). The binding of *lac* repressor and the catabolite gene activator protein to halogen-substituted analogues of poly (d(A-T)). Biochim. Biophys. Acta. *432*, 185–191.

Maniatis, T., Jaffrey, A. and Kleid, D. G. (1975). Nucleotide sequence of the rightward operator of phage. Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Martin, G. R., Epstein, C. J., Travis, B., Tucker, G., Yatziv, S., Martin, D. M., Clift, S. and Cohen, S. (1978) X-chromosome inactivation during differentiation of female teratocarcinoma stem cells in vitro. Nature *271*, 329–333.

McBurney, M. W. and Strutt, B. J. (1980). Genetic activity of X chromosomes in pluripotent female teratocarcinoma cells and their differentiated progeny. Cell *21*, 357–364.

McMaster, G. K. and Carmichael, G. G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agrose gels by using glyoxal and acridine orange. Proc. Nat. Acad. Sci. USA *74*, 4835–4838.

Miller, O. J., Schnedl, W., Allen, J. and Erlanger, B. F. (1974). 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature *251*, 636–637.

Mintz, B. and Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proc. Nat. Acad. Sci. USA *72*, 3585–3589.

Niwa, O. and Sugahara, T. (1981). 5-Azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation. Proc. Nat. Acad. Sci. USA *78*, 6290–6294.

Niwa, O., Decleve, A., Lieberman, M. and Kaplan, H. S. (1973). Adaptation of plaque assay methods to the in vitro quantitation of the radiation leukemia virus. J. Virol. *12*, 68–73.

Peries, J., Alves-Cardoso, E., Carlvet, M., Devons-Guillemin, M. C. and Lasneret, J. (1977). Lack of multiplication of ecotropic murine C-type viruses in mouse teratocarcinoma primitive cells. J. Nat. Cancer Inst. *59*, 463–465.

Razin, A. and Ceder, H. (1977). Distribution of 5-methylcytosine in chromatin. Proc. Nat. Acad. Sci. USA *74*, 2725–2728.

Rowe, W. P., Pugh, W. E. and Hartley, J. W. (1970). Plaque assay techniques for murine leukemia viruses. Virology *42*, 1136–1139.

Rowe, W. P., Hartley, J. W. and Besmer, T. (1972). Genetic mapping of a murine leukemia virus-inducing locus of AKR mice. Science *178*, 860–862.

Segal, S. and Khoury, G. (1979). Differentiation as a requirement for simian virus 40 gene expression in F-9 embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *76*, 5611–5615.

Sekikawa, K. and Levine, J. (1981). Isolation and characterization of polyoma host range mutants that replicate in nullipotential embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *78*, 1110–1104.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–518.

Speers, W. C., Gautsch, J. W. and Dixon, F. J. (1980). Silent infection of murine embryonal carcinoma cells by Moloney murine leukemia virus. Virology *105*, 241–244.

Stewart, C. L., Stuhlman, H., Jähner, D. and Jaenisch, R. (1982) *De novo* methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *79*, 4098–4102.

Strickland, S. and Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell *15*, 393–403.

Stuhlmann, H., Jähner, D. and Jaenisch, R. (1981). Infectivity and methylation of retrovirus genomes is correlated with expression in the animal. Cell *26*, 221–232.

Swartzendruber, D. E. and Lehman, J. M. (1975). Neoplastic differentiation: interaction of simian virus 40 and polyoma virus with murine teratocarcinoma cells in vitro. J. Cell Physiol. *85*, 179–188.

Swartzendruber, D. E., Friedrich, T. D. and Lehman, J. M. (1977). Resistance of teratocarcinoma stem cells to infection with simian virus 40: early events. J. Cell Physiol. *93*, 25–30.

Teich, N. M., Weiss, R. A., Martin G. R. and Lowy, D. R. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell *12*, 973–982.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Nat. Acad. Sci. USA *77*, 5201–5205.

Urano, Y., Kominami, R. and Muramatsu, S. (1980). The nucleotide sequence of the putative transcription initiation site of a cloned ribosomal RNA gene of the mouse. Nucl. Acids Res. *8*, 6043–6056.

van der Putten, H., Quint, W., Verma, I. M. and Berns, A. (1982). Moloney murine leukemia virus-induced tumors: recombinant provruses in active chromatin regions. Nucl. Acids Res. *10*, 577–592.

TAB 16

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 13724–13729, December 1997
Developmental Biology

Methylation of the minimal promoter of an embryonic globin gene silences transcription in primary erythroid cells

RAKESH SINGAL*, RICHARD FERRIS*, JANE A. LITTLE*†, SHU ZHEN WANG*, AND GORDON D. GINDER*†‡§

*Department of Medicine, Division of Medical Oncology; ‡Institute of Human Genetics; and †Cancer Center, University of Minnesota, Minneapolis, MN 55455-0362

Edited by Yuet Wai Kan, University of California, San Francisco, CA, and approved October 9, 1997 (received for review June 27, 1997)

ABSTRACT **Methylation of cytosines in the dinucleotide CpG has been shown to suppress transcription of a number of tissue-specific genes, yet the precise mechanism is not fully understood. The vertebrate globin genes were among the first examples in which an inverse correlation was shown between CpG methylation and transcription. We studied the methylation pattern of the 235-bp ρ-globin gene promoter in genomic DNA from primary chicken erythroid cells using the sodium bisulfite conversion technique and found all CpGs in the promoter to be methylated in erythroid cells from adult chickens in which the ρ-globin gene is silent but unmethylated in 5-day (primitive) embryonic red cells in which the gene is transcribed. To elucidate further the mechanism of methylation-induced silencing, an expression construct consisting of 235 bp of 5′ promoter sequence of the ρ-globin gene along with a strong 5′ erythroid enhancer driving a chloramphenicol acetyltransferase reporter gene, ρ-CAT, was transfected into primary avian erythroid cells derived from 5-day embryos. Methylation of just the 235-bp ρ-globin gene promoter fragment at every CpG resulted in a 20- to 30-fold inhibition of transcription, and this effect was not overridden by the presence of potent erythroid-specific enhancers. The ability of the 235-bp ρ-globin gene promoter to bind to a DNA Methyl Cytosine binding Protein Complex (MeCPC) was tested in electrophoretic mobility shift assays utilizing primary avian erythroid cell nuclear extract. The results were that fully methylated but not unmethylated 235-bp ρ-globin gene promoter fragment could compete efficiently for MeCPC binding. These results are a direct demonstration that site-specific methylation of a globin gene promoter at the exact CpGs that are methylated *in vivo* can silence transcription in homologous primary erythroid cells. Further, these data implicate binding of MeCPC to the promoter in the mechanism of silencing.**

Methylation of cytosine residues in the dinucleotide CpG is the most common postsynthetic eukaryotic DNA modification. Since the reports of an inverse correlation between DNA methylation and expression of vertebrate β-type globin genes (1–3), a large body of evidence relating DNA methylation to gene expression has accumulated (4, 5). At the same time, the absence of detectable DNA methylation in some eukaryotes such as *Drosophila* (6) and *Saccharomyces cerevisiae* (7) has raised doubts about its role in normal development and tissue-specific gene expression. However a study by Li *et al.* (8) showing abnormal development and embryonic lethality in transgenic mice expressing decreased but not completely absent DNA methyl transferase activity following knockout of the DNA methyl transferase gene lends strong support to a critical role for DNA methylation in developmental gene regulation. Recently, a similar critical function of DNA methylation in plant development has been demonstrated by Ronemus *et al.* (9). However, a direct demonstration of the role of DNA methylation in suppressing transcription of a specific gene during development in normal tissues has been lacking.

In the avian β-type globin cluster (5′-ρ-βH-β-ε-3′), silencing of the embryonic ρ-gene occurs concomitantly with activation of the adult β-gene on day 5 of embryonic development (10). In definitive embryonic and adult chicken erythroid cells, the embryonic globin genes are nontranscribed (11, 12), and a strong inverse correlation exists between site-specific DNA methylation and expression of the chicken β-type globin genes (13). Our laboratory has shown that the normally silent embryonic ρ-globin gene in red cells of anemic adult chicken can be transcriptionally activated by treatment with the DNA methyltransferase inhibitor, 5-azacytidine (14). However, the cause and effect relationship between methylation and stage-specific globin gene transcription in normal erythroid cells has not been fully elucidated. In this report, we describe (*i*) the *in vivo* methylation pattern of all CpG dinucleotides in a 235-bp minimal promoter of the embryonically expressed avian ρ-globin gene in 5-day primitive and definitive chicken erythroid cells, (*ii*) the transcription inhibitory effect of physiologically precise CpG methylation on the ρ-globin gene promoter despite the presence of strong 5′ and/or 3′ erythroid-specific enhancer elements in primary embryonic erythroid cells, and (*iii*) the demonstration of a DNA methyl binding protein complex that offers a possible mechanism of methylation-induced transcriptional silencing of the ρ-globin gene in normal erythroid cells.

MATERIALS AND METHODS

Vector Constructs. All transfection vectors were constructed in pUC 18. The vectors depicted in Fig. 2*A* were constructed by cloning a 2.2-kb ρ promoter region 5′ to a chloramphenicol acetyltransferase (CAT) reporter gene, pCAT Basic (Promega). Vectors in Fig. 2*B* were constructed by cloning a 3′ β/ε enhancer element 3′ into the ρ promoter containing CAT reporter gene constructs. Vectors in Fig. 2*C* were constructed by replacing the 2.2-kb ρ promoter with a 235-bp ρ-globin minimal promoter and cloning a DNA fragment containing 5′ hypersensitive sites 2 and 3 of the chicken β-globin cluster into the vector.

Methylation Reactions. Whole plasmids were methylated using Sss 1 methylase (New England BioLabs) following the manufacturer's protocol. The extent of methylation after each reaction was determined by digestion with *Msp*I and *Hpa*II. Region-specific methylation was carried out by excision of the fragment of interest by restriction digestion and gel isolation of the DNA fragment to be methylated. In each case, half of the

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9413724-6$2.00/0
PNAS is available online at http://www.pnas.org.

This paper was submitted directly (Track II) to the *Proceedings* office.
Abbreviations: MeCPC, methyl cytosine binding protein complex; CAT, chloramphenicol acetyltransferase.
§To whom reprint requests should be addressed at: Massey Cancer Center, Medical College of Virginia at Virginia Commonwealth University, 401 College St., P.O. Box 980037, Richmond, VA 23298-0037. e-mail: gginder@mcc1.mcc.vcu.edu.

154 264

DNA was methylated with Sss 1 methylase, and in a parallel control reaction the other half was incubated with methylase in the absence of *S*-adenosylmethionine (mock methylated) as a control. Methylated and mock-methylated DNA were religated into the vector DNA from which they were excised.

Transfection Assays. The protocol of Lieber *et al.* (15) was followed for primary avian erythroid cell transfections except that osmotic shock was in 300 mM NH_4Cl (pH 7.4) for 60 min at room temperature. Cells were cultured for 40–48 h in 10 ml of Leibovitz's culture medium. Cells were harvested, and cytoplasmic extracts were assayed for CAT activity by using a liquid scintillation assay kit (Promega) (16). After extraction, reactions were counted in a Beckman liquid scintillation counter. All transfections were carried out in at least three independent experiments and in triplicate, and results are expressed as mean values with standard errors of the mean for each construct. Controls for transfection efficiency were carried out using an Rous sarcoma virus promoter enhancer-driven β-galactosidase reporter gene construct.

Bisulfite Genomic Sequencing for Determination of DNA Methylation. After obtaining DNA from 5-day and adult chicken erythroid cells, bisulfite genomic sequencing was performed as described by Clark *et al.* (17) with the following modifications: (*i*) the bisulfite conversion reaction was carried out by incubating DNA with a 5 M bisulfite solution and 100 mM hydroquinone, pH 5.0, at 50°C for 4 h (18), and (*ii*) removal of free bisulfite was achieved by using a Qiaex II gel extraction kit (Qiagen). Sequencing of PCR-amplified product was performed using the dideoxy technique, [α-^{35}S]dATP internal labeling protocol using internal primers, primer F (5′-TGTAGGGGTGTTTTGTGTAAG-3′) and primer R (5′-CTATAAAAACACTCAAAACTTAAAAC-3′), constructed after taking into account the bisulfite conversion reaction. The sequence of the unmodified sense strand for which these primers were constructed is depicted in Fig. 1*B*. Improved sequencing results were, however, obtained using α-^{33}P-labeled ddNTP terminators.

Electrophoretic Mobility Shift Assays. HeLa cell and erythroid cell nuclear extracts were prepared according to the method of Dignam *et al.* (19). CG11 oligonucleotide (135 bp) containing 20 *Hha*I sites and 7 *Hpa*II sites has been described (20). The rho 235 promoter fragment was obtained by restriction digestion of the plasmid rho 235-CAT. Probes were labeled using the Klenow fragment of DNA polymerase I and [α-^{32}P]dATP. The electrophoretic mobility shift assay was performed as described (20). *Micrococcus lysodeikticus* DNA digested with *Sau*3aI was used as nonspecific competitor in all reactions.

RESULTS

Methylation Analysis of the 235-bp Minimal Promoter of the ρ-Globin Gene in Genomic DNA. Though an inverse correlation between DNA methylation and expression of avian β-type globin genes (1, 13) has been reported, the methylation pattern of only a limited number of CpG dinucleotides has been described. Earlier techniques have depended on phosphodiester bond cleavage adjacent to cytosine residues using either chemical cleavage by hydrazine in Maxam and Gilbert sequencing reactions (21), which cleaves cytosine but not 5-methylcytosine, or methylation-sensitive restriction enzymes. As 5-methylcytosine is not cleaved in Maxam and Gilbert sequencing reactions, it is detected as a gap in a sequencing ladder, which may be difficult to interpret. Methylation-sensitive restriction enzyme-based techniques can only analyze a limited number of potentially methylated sites (22, 23). Recently, a technique has been described in which bisulfite-induced modification of genomic DNA was carried out under conditions that convert cytosine to uracil but that do not affect 5-methylcytosine. PCR amplification and sequencing of the modified DNA results in a detected sequence pattern showing conversion of all unmethylated cytosines to thymidine (since uracil is read by *Taq* polymerase as thymidine), whereas 5-methylcytosines remain as cytosines (17, 24). We used this technique to study the methylation pattern of CpG dinucleotides in the 235-bp minimal promoter of ρ-globin gene from erythroid cell genomic DNA of 5-day embryos and adult chickens. Previous reports have shown this promoter is capable of supporting high level, developmental specific transcription in primary avian erythroid cells (25). After bisulfite treatment, attempts to amplify a product larger than 600 bp were unsuccessful, as we were unable to detect a discrete product similar to the experience of Selker *et al.* (26). We were, however, able to obtain a discrete PCR product of 467 bp. Notably, similar

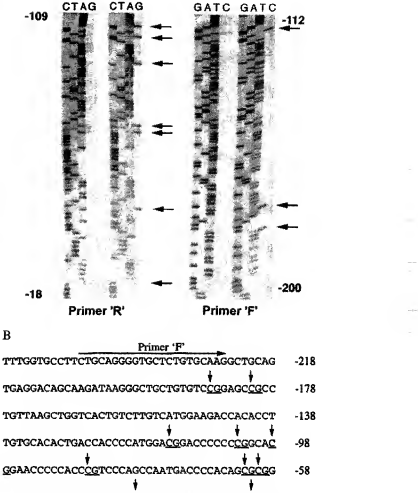


FIG. 1. *In vivo* methylation of CpG dinucleotides of the rho promoter in 5-day and adult chicken erythroid cells using the bisulfite conversion technique. *Arrows* indicate methylated cytosines. Positions indicated are relative to the transcription start site (*A*). Cytosines that are not associated with CpG dinucleotides (sequence shown in *B*) have all been converted to thymidines in both 5-day and adult erythroid cells. (*B*) ρ-Globin gene promoter sequence. *Arrows* indicate methylated cytosines that are clearly seen in *A* (data not shown for CpG dinucleotide at position −15). Primers R and F indicate the sequence of ρ-globin gene promoter used for designing internal primers and for sequencing.

reactions using genomic DNA not treated with bisulfite resulted in only a broad smear in ethidium-stained agarose gels. Dideoxy sequencing of the purified PCR product was carried out using initially [α-^{35}S]dATP internal labeling protocol and subsequently with α-^{35}P-labeled ddNTP terminators. As shown in Fig. 1*A* and in data not shown, sequencing of the PCR product showed the presence of cytosine residues (indicated by *arrow*) only in PCR products obtained from adult erythroid cell DNA. All unmethylated cytosine residues in the original sequence were converted to thymidines, demonstrating the completeness of the bisulfite conversion.

As summarized in Fig. 1*B*, cytosine residues in all CpG dinucleotides in the ρ-gene promoter were found to be methylated in the 235-bp minimal promoter of the ρ-globin gene in DNA from definitive adult chicken erythroid cells. In contrast, none of the cytosine residues were methylated in CpG dinucleotides from the same region in primitive 5-day embryonic erythroid cells.

Methylation of Whole Plasmid Containing the ρ-Gene 2.2-kb Upstream Promoter Driving the CAT Reporter Gene Suppresses Transcription. We have found previously that methylation of a genomic ρ-globin gene construct containing 2.2 kb of upstream promoter sequences blocked transcription in stably transfected murine erythroleukemia cells (27). To study the cause and effect relationship between methylation and stage-specific globin gene transcription in normal erythroid cells, an expression construct consisting of 2.2 kb of ρ-globin gene promoter driving a CAT reporter gene was used to transfect primary erythroid cells from 5-day chicken embryos in a transient transfection assay system. Since all CpG dinucleotides were found to be methylated in the 235-bp minimal promoter from adult erythroid cell genomic DNA, methylation experiments were carried out using the enzyme Sss 1 methylase, which methylates all CpG dinucleotides. Methylation of the whole plasmid resulted in almost no detectable promoter activity. To assess the combination of individual components, methylation of the 2.2-kb promoter was carried out separately, and the product was religated into the unmethylated plasmid and transfected. Methylation of the 2.2-kb promoter resulted in an approximately 30-fold reduction in promoter activity compared with that from a mock-methylated control. A similar experiment, in which the CAT reporter gene alone was methylated and religated into the otherwise unmethylated plasmid construct and transfected, interestingly showed similar suppression of transcription compared with mock-methylated control (Fig. 2*A*).

Methylation of Whole Plasmid or 2.2-kb Upstream Promoter but Not the 3′ β/ε Enhancer Suppresses Transcription. To see if the presence of a strong erythroid enhancer would override methylation-induced transcriptional repression, a 480-bp fragment containing chicken β-globin cluster the β/ε 3′ enhancer was ligated into the ρ-CAT plasmid described in Fig. 2*A*. Methylation of whole plasmid or the 2.2-kb ρ-promoter alone resulted in marked reduction in promoter activity. In contrast, a similar experiment in which the 3′ β/ε enhancer alone was methylated showed no significant difference as compared with mock-methylated control (Fig. 2*B*).

Methylation of the 235-bp ρ-Gene Promoter but Not the pUC-18 Vector Suppresses Transcription Despite the Strong 5′ Enhancer (Hypersensitive Sites 2 and 3). As described (25), the major activity of the ρ-gene promoter seems to be contained within sequences lying between −246 and the transcription start site. A construct was therefore made consisting of a ρ-gene promoter (base pairs 0 to −235) driving a CAT reporter gene in which a DNA fragment containing 5′ hypersensitive sites 2 and 3 was cloned. Methylation of the 235-bp ρ-gene promoter resulted in an approximately 20-fold decrease in promoter activity even in the presence of a strong 5′ erythroid enhancer (hypersensitive sites 2 and 3) (Fig. 2*C*). These 5′ hypersensitive sites have been demonstrated to have strong erythroid-specific enhancer activity in transfection assays and transgenic mice (28, 29, 43). In contrast, complete methylation of the pUC-18 vector backbone, which contains 157 CpG dinucleotides, had a negligible effect on transcription from the ρ-gene promoter (Fig. 2*C*).

M-rho 235 Competes for MeCPC Binding. The reported mechanisms of methylation-induced transcriptional repression seem mainly to be either by directly preventing the binding of transcription factors to the promoter or indirectly through proteins that bind preferentially to methylated DNA (30). Among transcription factors known to be sensitive to methyl-CpG, none seems to have a canonical binding site in the ρ-gene 235-bp promoter fragment (25). The known proteins and complexes that bind preferentially to methylated DNA include methylCpG binding protein 1 (MeCP-1) (20), methylCpG binding protein 2 (MeCP-2) (31), and methylated DNA-binding protein 2 (MDBP2) (32). MDBP2 has been implicated in the transcriptional repression of the vitellogenin gene, which could be overcome by estradiol stimulation (33). Subsequently purified MDBP2 was found to be histone H1 (34). Though some investigators have proposed preferential binding of histone H1 to methylated DNA and thereby implicated it in methylation-mediated transcriptional repression (35, 36), others have found this not to be the case (37). MeCP-1 binds *in vitro* to DNA containing at least 12 symmetrically methylated CpGs (20), whereas MeCP-2 can bind to a single methylated CpG (31). Though MeCP-2 seems to be essential for mouse development and despite earlier findings that it did not affect transcription of methylated promoter specifically (38), it has been shown recently to block transcription in cell lines (39).

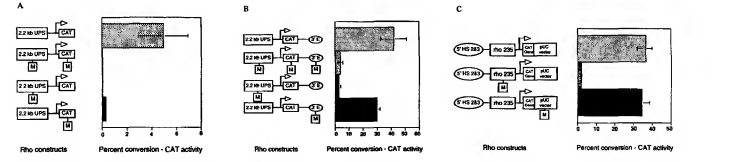


FIG. 2. Levels of ρ-globin gene promoter activity in primary chicken erythroid cells transiently transfected with the methylated or mock-methylated ρ-globin gene constructs depicted. Complete methylation of all CpGs in a particular region of a given construct is represented by attached M. Each experiment was carried out in triplicate, and levels of CAT conversion percentages presented represent the mean values and standard deviations. Methylation of the extended 2.2-kb upstream promoter resulted in significant reduction in promoter activity in the absence (*A*) or presence (*B*) of strong erythroid 3′ β/ε enhancer. Methylation of the 235-bp ρ-globin gene promoter also resulted in significant reduction in promoter activity even in the presence of strong 5′ enhancer, which includes erythroid-specific hypersensitive sites 2 and 3, whereas methylation of the pUC vector alone had no effect (*C*).

MeCP-1 has been implicated in methylation-mediated transcriptional repression of several genes including the human α-globin gene in heterologous cultured cells (40, 41). F9 cells, which contain low levels of MeCP-1, cannot efficiently repress those methylated gene promoters (41).

Because the ρ-gene 235-bp promoter sequence has multiple CpGs that were found to be completely methylated in genomic DNA from adult chicken erythroid cells in which the gene is repressed, we turned our attention toward MeCP-1 as a possible mediator of transcriptional repression in the primary erythroid cell transfection assay. End-labeled methylated and mock-methylated CG11 probes (M-CG11 and CG-11, respectively) were mixed with HeLa cell nuclear extract and then subjected to electrophoretic mobility shift assays to detect MeCP-1-like complex as described (20). On autoradiography, a complex was observed with M-CG11, but none was detected with CG11 as shown in Fig. 4. Similar to the findings reported earlier (20, 40), this complex was effectively competed by an excess of cold M-CG11 as well as M-rho 235 but not by excess of cold, unmethylated CG11 or rho 235 (Fig. 3).

M-CG11 Competes for MeCP-1-Like Complex Binding by M-rho 235. Because M-rho 235 effectively competed for MeCP-1 binding by M-CG11, the possibility that it could directly form a complex with MeCP-1 was examined. In a reverse experiment, end-labeled methylated and mock-methylated rho 235 probes were used in band shift assays with HeLa cell nuclear extract. A complex was observed with M-rho 235, but none was detected with (mock-methylated) rho 235. This complex was effectively competed by an excess of unlabeled M-rho 235 but not by an excess of unlabeled rho 235. Competition of this complex with M-CG11 and CG11 showed effective competition by M-CG11 but not by (mock-methylated) CG11 (data not shown).

MeCPC Is Present in Nuclear Extracts from Primary Erythroid Cells and a Similar Complex Binds to Methylated rho 235. Having demonstrated the binding of MeCPC to methylated rho 235 using HeLa cell nuclear extract, its presence in the primary chicken erythroid cells used in the transient transfection functional assays was investigated. End-labeled methylated and mock-methylated CG11 probes were mixed with erythroid cell nuclear extracts derived from 8- and 11-day chicken embryos. Electrophoretic mobility shift assays showed the presence of a complex with similar mobility as seen with HeLa cell nuclear extracts, with M-CG11 but not with CG11, as shown in Fig. 4*A*.

Previous studies of MeCP-1 have relied on competition between methylated promoter sequences and methylated CG11 probe to demonstrate its interaction with the promoter under study. To extend these findings, we investigated directly the ability of methylated rho 235 minimal promoter to bind to MeCP-1-like methylation-dependent complex using homologous primary avian erythroid cell nuclear extracts.

As illustrated in Fig. 4*B*, the formation of such a complex with the rho 235-bp promoter sequence was detected in 5-day (primitive) and 11-day (definitive) avian erythroid cell nuclear extracts. This methylation-dependent complex was easily competed by an excess of unlabeled methylated rho 235 but not by an excess of unlabeled mock-methylated rho 235.

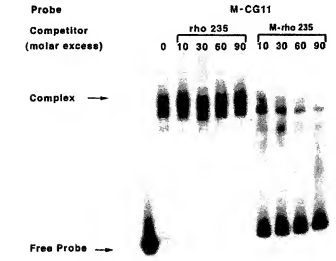


FIG. 3. Mobility shift assay for MeCPC binding to methylated CG11 probe (M-CG11). 0.6 ng of methylated CG11 end-labeled probes were mixed with 6 μg of HeLa cell nuclear extract and varying amounts of mock-methylated or methylated rho 235 promoter DNA fragment. The reaction mixture was electrophoresed in a 1.5% agarose gel, and complexes were analyzed by autoradiography. This autoradiogram illustrates the effective competition of MeCPC binding to M-CG11 probe by methylated rho 235 but not by mock-methylated rho 235 (M-rho 235).

DISCUSSION

The principal findings in this report are as follows. Every cytosine within a CpG dinucleotide in a 235-bp region of the embryonic ρ-globin gene promoter is methylated in normal adult (definitive) erythroid cells in which the gene is silent and completely unmethylated in 5-day (primitive) erythroid cells in which the gene is actively transcribed. Previous studies in our laboratory have shown extensive CpG methylation in a 4.6-kb fragment that includes 1.5 kb of 5′ and 3′ of ρ-gene flanking sequences in definitive erythroid cells (14). *In vitro* methylation of those same CpG sites in the 235-bp ρ-gene promoter fragment markedly reduces the promoter activity in a transient transfection assay in *ex vivo* primary avian erythroid cells, despite the presence of a strong erythroid-specific enhancer. The silencing effect of methylation is not overridden even in the nuclear environment of 5-day erythrocytes, which contain all the nuclear factors necessary for high level transcription of the ρ-gene. Finally, we show that a methyl CpG binding protein complex, which behaves like MeCP-1 (20), binds to methylated but not unmethylated ρ-promoter sequences.

Numerous reports have shown the ability of promoter DNA methylation to inhibit transcription of a wide variety of genes, and in some cases such methylation corresponds to the inactive state of the gene under study *in vivo* (reviewed in refs. 4 and 5). In the case of globin genes, inhibition of expression of the human γ-globin gene in nonerythroid murine L-cells was observed when all CpGs in the promoter were methylated in a stable transfection assay (42). More recently, it was shown that methylation of all CpGs in the human α-globin gene could block transcription in transfected nonerythroid HeLa cells and in cell-free *in vitro* transcription assays derived from the same cell type (40, 41). In neither of these studies was the *in vivo* methylation pattern of all CpGs of the specific globin gene in question reported, nor was the pattern of methylation or its effect on developmental stage-specific expression in primary erythroid cells tested.

In this report we show that methylation of the embryonic ρ-globin gene promoter is capable of attenuating transcription in primary erythroid cells even in the presence of strong erythroid-specific enhancers that we have shown (25, 43) to be capable of overcoming the stage-specific silencing effects of the ρ-gene promoter in primary erythroid cells. However, methylation of the 3′ β/ε erythroid enhancer or the pUC-18 vector backbone had no effect, demonstrating that the blocking action of CpG methylation on ρ-promoter driven transcription resides in the sequences of the transcription unit. Of interest, the level of transcription for the methylated rho 235-bp promoter was significantly higher in the presence of the strong erythroid enhancer. Quantitation of the increased transcription levels conferred by either the 3′ β/ε on 5′

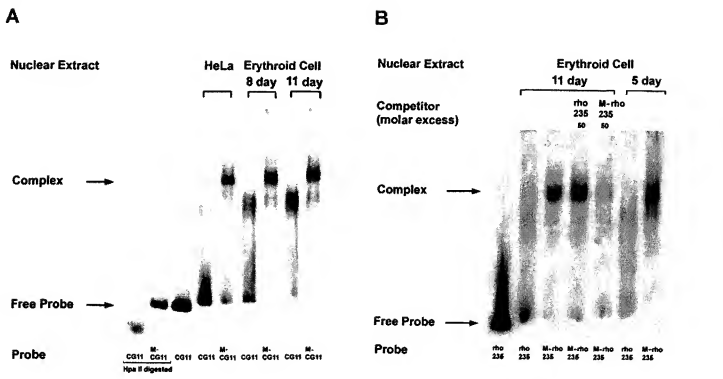


FIG. 4. Mobility shift assay for the presence of MeCPC in chicken erythroid cell or control HeLa cell nuclear extracts. End-labeled, methylated CG11 probe (M-CG11) but not mock-methylated CG11 (CG11) forms a complex with erythroid cell nuclear extract that has similar mobility as seen with HeLa cell nuclear extract (*A*). Similar results were seen in 5-day chicken erythroid cell nuclear extract (data not shown). (*B*) Binding of MeCPC in chicken erythroid cell nuclear extract to methylated rho 235 (M-rho 235) probe. End-labeled M-rho 235 but not mock-methylated rho 235 forms a complex with erythroid cell nuclear extract. This complex was competed effectively by an excess of unlabeled methylated but not mock-methylated rho 235.

hypersensitive site (HS) enhancers indicated that the magnitude of increase (10–20-fold) was approximately the same for methylated vs. unmethylated promoter constructs in concomitant transfection assays (data not shown). Thus, even though methylation decreased the frequency of productive transcriptional complex formation by one to two orders of magnitude, the two erythroid enhancers tested could still markedly increase that frequency in homologous primary erythroid cells. These results support the concept that tissue-specific enhancers can at least partially override repressive influences on promoters, at least as assayed in transient transfection systems.

Previous studies have shown the ability of the methylated DNA binding protein complex, MeCP-1, to block transcription of the human α-globin gene promoter even in the presence of the strong simian virus 40 enhancer (40, 41). Here, we show that a similar methylated DNA binding complex forms efficiently with the methylated *p*-gene promoter sequence and that the complex can be detected using nuclear extracts from the same primary avian erythroid cells in which methylation-mediated transcriptional inhibition was demonstrated. Thus, these results support a role for MeCP-1 or a similar complex in developmental silencing of embryonic globin genes in normal erythropoiesis. Determination of whether the complex formed between the methylated *p*-promoter and nuclear extracts from primary avian erythroid cells is the same as MeCP-1 will await full characterization of both of these large protein–DNA complexes. Direct evidence that this complex blocks transcription in this system will require additional studies, which are underway.

Alternatively, it is possible that methylation of the *p*-gene promoter directly blocks binding of one or more specific transcriptional activating DNA binding factors. This seems unlikely as none of the sequence recognition sites for the known activating factors that bind to the *p*-promoter contain CpGs in their core recognition sequences (25). Also, such a localized sequence-specific effect was not observed for the human γ-globin or α-globin genes (40, 42). Further support for some type of blocking effect on transcription complex formation, promoter clearance, or transcription elongation is provided by the experiment in which methylation of the CAT reporter cassette sequences only (which also contain multiple CpGs) inhibited transcription. Quantitation of the CAT reporter assay showed that methylation of the CAT cassette inhibited expression by 5-fold less than methylating the *p*-promoter sequences. This result, coupled with the finding that methylation of the entire expression plasmid did not inhibit CAT expression more than methylation of the 235-bp *p*-promoter alone, suggests that the promoter methylation observed in definitive erythroid cells is a major mediator of methylation-induced silencing of this gene *in vivo*. Although binding of MeCP-1 by the methylated CAT sequences has not yet been demonstrated, this seems a likely mechanism for the observed effect. Regardless, this result does support the notion that the repressive effect of CpG methylation need not necessarily be mediated through the promoter.

The results presented here also support the notion that DNA methylation inhibits globin gene expression in a stage-specific fashion, at least in part by a direct interference with transcription initiation, elongation, or promoter clearance rather than by inducing an inactive chromatin structure, as transient transfection of nonreplicating cells does not result in intact chromatin structure. It is still possible that additional effects of CpG methylation downstream of the promoter and in the context of intact chromatin may not be detected in the transient transfection assays used in this report and previous studies (40, 41). However, a direct effect on the transcriptional machinery is most consistent with the evidence from several labs, including our own, suggesting that the entire avian

β-globin gene cluster is in an open chromatin domain in erythroid cells of all developmental stages (44–46).

The potent transcriptional silencing effect of physiologically accurate DNA methylation of the ρ-gene promoter even in the nuclear environment of transcriptionally active and stage-specific primary erythroid cells suggests that methylation plays a dominant role in maintaining the high level of repression of embryonic globin gene transcription in definitive or adult erythroid cells. By analogy, such a strong suppressive effect of DNA methylation suggests a role for active demethylation in relieving repression of CpG-rich promoters that are highly methylated in cells containing abundant MeCP-1 or like complexes. Recently, evidence for such sequence-specific demethylating activity has been reported (47).

Strong repression of transcription from the ρ-globin gene during the embryonic developmental switch from primitive to definitive erythroid cell lineages is not likely a prerequisite for survival. Nonetheless, it seems feasible that mechanisms analogous to those described here may control genes that are critical for embryonic differentiation and viability, as evidenced by the results obtained in DNA methylase or methyl binding protein-deficient mouse embryos (8) and embryonic stem cells (48). The results presented in this report are direct evidence of a transcriptional inhibitory role for DNA methylation in the silencing of a specific developmentally regulated gene in homologous primary embryonic cells.

We thank Joan Boyes, National Institutes of Health, for providing plasmids CG11 and *M. lysodeikticus* and for technical advice; we also thank Vivian Bardwell and Kathleen Conklin for critical reading of the manuscript. This work was supported by National Institutes of Health Grant R-37DK29902 and the Masonic Cancer Center Professorship Fund.

1. McGhee, J. D. & Ginder, G. D. (1979) *Nature (London)* **280**, 418–420.
2. Shen, C. K. J. & Maniatis, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6634–6638.
3. Van der Ploeg, L. H. T. & Flavell, R. A. (1980) *Cell* **19**, 947–958.
4. Razin, A. & Cedar, H. (1991) *Microbiol. Rev.* **55**, 451–458.
5. Bird, A. (1992) *Cell* **70**, 5–8.
6. Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. (1982) *FEBS Lett.* **146**, 148–152.
7. Proffitt, J. H., Davie, J. R., Swinton, D. & Hattman, S. (1984) *Mol. Cell. Biol.* **4**, 985–988.
8. Li, E., Bestor, T. H. & Jaenisch, R. (1992) *Cell* **69**, 915–926.
9. Ronemus, M. J., Galbiati, M., Ticknor, C., Chen, J. & Dellaporta, S. L. (1996) *Science* **273**, 654–657.
10. Brown, J. L. & Ingram, V. M. (1974) *J. Biol. Chem.* **249**, 3960–3972.
11. Groudine, M., Peretz, M. & Weintraub, H. (1981) *Mol. Cell. Biol.* **1**, 281–288.
12. Landes, G. M. & Martinson, H. G. (1982) *J. Biol. Chem.* **257**, 11002–11007.
13. Ginder, G. D. & McGhee, J. D. (1981) in *Globin Gene Expression and Hematopoietic Differentiation*, eds. Stamatoyannopoulos, G. & Nienhuis, A. W. (Liss, New York), pp. 191–201.
14. Ginder, G. D., Whitters, M. J. & Pohlman, J. K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3954–3958.
15. Lieber, M. R., Hesse, J. E., Nickol, J. M. & Felsenfeld, G. (1987) *J. Cell Biol.* **105**, 1055–1065.
16. Gustafson, K. S. & Ginder, G. D. (1996) *J. Biol. Chem.* **271**, 20035–20046.
17. Clark, S. J., Harrison, J., Paul, C. L. & Frommer, M. (1994) *Nucleic Acids Res.* **22**, 2990–2997.
18. Raizis, A. M., Schmitt, F. & Jost, J.-P. (1995) *Anal. Chem.* **226**, 161–166.
19. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
20. Meehan, R. R., Lewis, J. D., McKay, S., Kleiner, E. L. & Bird, A. P. (1989) *Cell* **58**, 499–507.
21. Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
22. Singer, J., Roberts-Ems, J. & Riggs, A. D. (1979) *Science* **203**, 1019–1021.
23. Singer-Sam, J., Grant, M., LeBon, J. M., Okuyama, K., Chapman, V., Monk, M. & Riggs, A. D. (1990) *Mol. Cell. Biol.* **10**, 4987–4989.
24. Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L. & Paul, C. L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831.
25. Minie, M. E., Kimura, T. & Felsenfeld, G. (1992) *Development* **115**, 1149–1164.
26. Selker, E. U., Fritz, D. Y. & Singer, M. J. (1993) *Science* **262**, 1724–1728.
27. Wandersee, N. (1996) Ph.D. thesis (Univ. of Iowa, Iowa City).
28. Reitman, M., Lee, E. & Westphal, H. (1995) *Nucleic Acids Res.* **23**, 1790–1794.
29. Mason, M. M., Lee, E., Westphal, H. & Reitman, M. (1995) *Mol. Cell. Biol.* **15**, 407–414.
30. Tate, P. H. & Bird, A. P. (1993) *Curr. Opin. Gen. Dev.* **3**, 226–231.
31. Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurer-Fogy, I., Jeppesen, P., Klein, F. & Bird, A. (1992) *Cell* **69**, 905–914.
32. Pawlak, A., Bryans, M. & Jost, J.-P. (1991) *Nucleic Acids Res.* **19**, 1029–1034.
33. Jost, J. P., Saluz, H. P. & Pawlak, A. (1991) *Nucleic Acids Res.* **19**, 5771–5775.
34. Jost, J. P. & Hofsteenge, J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9499–9503.
35. Levine, A., Yeivin, A., Ben-Asher, E., Aloni, Y. & Razin, A. (1993) *J. Biol. Chem.* **268**, 21754–21759.
36. Johnson, C., Goddard, J. & Adams, R. (1995) *Biochem. J.* **305**, 791–798.
37. Bird, A., Tate, P., Nan, X., Campoy, J., Meehan, R., Cross, S., Tweedie, S., Charlton, J. & Macload, D. (1995) *J. Cell Sci.* **19**, Suppl., 37–39.
38. Meehan, R. R., Lewis, J. D. & Bird, A. P. (1992) *Nucleic Acids Res.* **20**, 5085–5092.
39. Nan, X., Campoy, F. J. & Bird, A. (1997) *Cell* **88**, 471–481.
40. Boyes, J. & Bird, A. (1992) *EMBO J.* **11**, 327–333.
41. Boyes, J. & Bird, A. (1991) *Cell* **64**, 1123–1134.
42. Busslinger, M., Hurst, J. & Flavell, R. A. (1983) *Cell* **34**, 197–206.
43. Wandersee, N. J., Ferris, R. C. & Ginder, G. D. (1996) *Mol. Cell. Biol.* **16**, 236–246.
44. Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. & Weintraub, H. (1980) *Cell* **20**, 451.
45. Wood, W. I. & Felsenfeld, G. (1982) *J. Biol. Chem.* **257**, 7730–7736.
46. Brotherton, T. W., Reneker, J. & Ginder, G. D. (1990) *Nucleic Acids Res.* **18**, 2011–2016.
47. Weiss, A., Keshet, I., Razin, A. & Cedar, H. (1996) *Cell* **86**, 709–718.
48. Lei, H., Oh, S. P., Okano, M., Jüttermann, R., Goss, K. A., Jaenisch, R. & Li, E. (1996) *Development* **122**, 3195–3205.

TAB 17

This material may be protected by Copyright law (Title 17 U.S. Code)

ELSEVIER

Gene 261 (2000) 181–188

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Retroviruses in foreign species and the problem of provirus silencing ☆

J. Svoboda[a,*], J. Hejnar[a], J. Geryk[a], D. Elleder[a], Z. Vernerová[b]

[a]*Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n. 2, 166 37 Prague 6, Czech Republic*
[b]*Department of Pathology, First Faculty of Medicine, Charles University, Prague, Czech Republic*

Received 29 May 2000; received in revised form 18 August 2000; accepted 5 October 2000
Received by G. Bernardi

Abstract

Retroviruses are known to integrate in the host cell genome as proviruses, and therefore they are prone to cell-mediated control at the transcriptional and posttranscriptional levels. This plays an important role especially after retrovirus heterotransmission to foreign species, but also to differentiated cells. In addition to host cell-mediated blocks in provirus expression, also so far undefined host specificities, deciding upon the pathogenic manifestation of retrovirus heterotransmission, are in play. In this respect, we discuss especially the occurrence of wasting disease and immunodeficiency syndrome, which we established also in avian species using avian leukosis virus subgroup C (ALV-C) inoculated in mid-embryogenesis in duck or chicken embryos. The problem of provirus downregulation in foreign species or in differentiated cells has been in the recent years approached experimentally. From a series of observations it became apparent that provirus downregulation is mediated by its methylation, especially in the region of proviral enhancer-promoter located in long terminal repeats (LTR). Several strategies have been devised in order to protect the provirus from methylation using LTR modification and/or introducing in the LTR sequence motifs acting as antimethylation tags. In such a way the expression of retroviruses and vectors in foreign species, as well as in differentiated cells, has been significantly improved. The complexity of the mechanisms involved in provirus downregulation and further possibilities to modulate it are discussed. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Retrovirus heterotransmission; Retroviral vector: Methylation; CpG; Long terminal repeat(s); Histone deacetylase

1. Introduction

The topic of this review is focused on selected aspects of retrovirus heterotransmission, but also touches the problem of retroviral genome silencing in differentiated and other types of cells. In case of heterotransmission, many host cell factors are in play, which decide whether or to what degree the retroviral genome will be expressed and what pathogenic consequences may be triggered. As is known generally, the retroviral genome becomes integrated in the cell genome as a provirus and, therefore, it is not surprising that it is highly influenced by the host-cell gene-regulation machinery. In discussing such downregulation of the provirus we are in fact dealing with post-integration blocks in provirus expression, the nature of which is epigenetic and mediated by new host cell factors. The unusual cell milieu, in concert with which the virus has not been evolving, can be lacking some factors like those enabling viral RNA export from the nucleus, or provide unusual factors like those changing the viral RNA splicing. However, of main importance is cell transcriptional regulation, which in many cases leads to provirus silencing. It is, therefore, not surprising that in phylogenetically distant host cells the provirus can integrate, but in many cases does not produce an infectious progeny. We call such host cells non-permissive, in contrast to permissive cells where formation of infectious virions takes place.

Non-permissivenes to retroviral infection has been for a

Abbreviations: ALV, avian leukosis virus(es); ALV-B, ALV-C, ALV-D, avian leukosis virus subgroup B, C, and D, respectively; *aprt*, adenosine-phosphoribosyltransferase gene; β-*geo*, fused β-galaetosidase and neomycin resistance gene; BLV, bovine leukemia virus; CAT, chloramphenieol acetyltransferase; cHS4, chicken hypersensitive site 4; EC, embryonic carcinoma; GFP, green fluorescence protein; HDAC, histone deacetylase; HIV, human immunodeficiency virus; IFN-SAR, interferon scaffold attachment region; LCR, locus control region(s); LTR, long terminal repeat(s); MeCP, methyl-CpG-binding protein; MEL, murine erythroleukemia; MLV, murine leukemia virus(es); MSV, murine sarcoma virus; NCR, negative control region; *neo*, neomycin resistance gene; PR RSV, Prague strain of RSV; RSV, Rous sarcoma virus; SIV, simian immunodeficiency virus; SIVcpz, chimpanzee SIV; SIVsm, sooty mangabey SIV; TSA, trichostatin; X-MLV, xenotropic MLV

☆ Extended version of the lecture "Modification of retrovirus pathogenicity by transspecies transfer" presented at the workshop "Microbial Variation and Evolution" at Ischia, Italy (organizers G. Bernardi and B. Fantini).

* Corresponding author. Tel.: +42-2-243-10-238; fax: +42-2-243-10-955.

E-mail address: svoboda@img.cas.cz (J. Svoboda).

0378-1119/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(00)00481-9

long time ascribed to foreign species host cells. Interestingly enough, non-permissiveness to retroviral infection marks also some cells of the virus species origin. This has been recognized especially in cases of cultured differentiated cells. In such a way, surmounting hurdles negatively influencing provirus expression became of importance for an efficient application of retroviral vectors for gene therapy of developmentally committed stem cells and differentiated cells. Therefore, we also discuss the provirus fate in some defined cases of differentiated cells that originate from the same species as the retrovirus. There are common features shared among all these situations, pointing to a role of cell-mediated gene silencing as an important factor that superimposes upon provirus expression.

2. Trans-class retrovirus heterotransmission

This extreme situation was first achieved in the case of chicken Rous sarcoma virus (RSV) strains transmitted first to rodents and later also to other mammalian species, including monkeys (rev. Svoboda, 1986). RSV genome expression in mammalian cells is governed by a series of factors, especially by flanking DNA sequences, their richness in GC (Fincham and Wyke, 1991; Rynditch et al., 1991), but also posttranscriptional steps are involved, which are blocked in mammalian cells (rev. Svoboda, 1998). More recent progress indicates that of importance are RSV LTR, which are prone to methylation in mammalian cells.

It was shown experimentally that a reporter gene driven by in vitro methylated RSV LTR is more efficiently suppressed in mammalian cells as compared to chicken cells (Hejnar et al., 1999). In spite of the fact that there are 16 CpGs in the Prague strain of Rous sarcoma virus (PR RSV) LTR and multiple CpGs are present in leader sequences, methylation of one CCGG *Hpa*II site, located downstream of the promoter region but close to the single provirus transcriptional start, was sufficient to produce dramatic reporter downregulation in mammalian cell lines. It should be noted that these experiments were done using transient transfection of methylated proviral DNAs and unmethylated controls. Expression of unmethylated LTR was comparable in both avian and non-permissive mammalian cells, suggesting that both types of cells harbor sufficient transcriptional machinery required by RSV LTR. Silencing of RSV proviruses is therefore a post-integration event.

The significance of LTR methylation for provirus downregulation has been recently approached using two strategies. The first one, schematically shown in Fig. 1, implies insertion of four canonical Sp1 binding sites in the RSV LTR enhancer region using *Eco*RI sites. Such reconstruction has been documented and discussed in detail (Machoň et al., 1998), and it was found that Sp1 insertion significantly increases LTR-driven chloramphenicol acetyltransferase (CAT) reporter gene expression, especially in hamster cells using both transient and stable transfection assays. This is in agreement with findings revealing that Sp1 binding sites represent the critical part of sequences acting as antimethylation tags, as was documented clearly in the case of a CpG island containing 1.7 kb DNA located in front of the adenosine-phosphoribosyltransferase (*aprt*) gene (Mummaneni et al., 1993; Brandeis et al., 1994; Macleod et al., 1994). Acquisition of an Sp1 binding site by mutation in the LTR enhancer region was shown previously to activate MLV (murine leukemia virus) transcription in embryonic carcinoma (EC) cells (Prince and

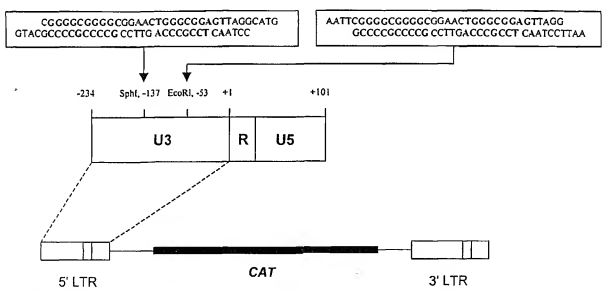


Fig. 1. Scheme of the CAT reporter vector employed for insertion of Sp1 sites into LTR sequence and transfection experiments. Sites of insertion (*Sph*I and *Eco*RI) are denoted by arrows. Sp1 binding sites within the inserted sequences are in bold.

Rigby, 1991). In a similar way, transcriptional activation of human endogenous retrovirus (HERV-H) was accomplished by Sp1 occurrence in either LTR enhancer (Nelson et al., 1996) or promoter (Sjøttem et al., 1996; Anderssen et al., 1997). The second approach is based on the incorporation of the mouse *aprt* gene CpG island immediately upstream to the RSV LTR-driven fused β-galactosidase and neomycin resistance (β-*geo)* proviral reporter (see Fig. 2). After transfection of hamster cells and selection for neomycin resistance, such construction ensured stable transcriptional activity in a reasonable number of cell clones, whereas unprotected proviral reporters are inactivated. This effect can be attributed to the antimethylation protection of the CpG island for two reasons; first, the active proviruses remained unmethylated within the 5′ LTR as evidenced by the bisulphite sequencing technique. second, the transcription efficiency of the RSV LTR is not increased in the presence of the CpG island.

Thus, both approaches led to the conclusion that proviral LTR could be protected from DNA methylation in the foreign species host. Optimization of these protective strategies might open the way to construct improved RSV-based vectors for gene transfer, more suitable for expression in mammalian cells and without the risk of infectious retroviral progeny. Retroviral vector producer cells (helper cells) are an important object for such protection as well. These cells have been shown genetically unstable due to the methylation of integrated helper proviral constructions. Designing a helper virus to overcome cellular DNA methylation may therefore improve vector production (Young et al., 2000).

There exists also another way of trans-species retrovirus transmission utilizing xenotropic murine leukemia viruses (X-MLV) (rev. Levy, 1978). It is interesting that these viruses can replicate in some avian cells such as duck cells, but not in others like chicken cells (Levy, 1977). X-MLV can provide envelope components to pseudotype RSV virions and such pseudotypes transform and replicate in duck but not in chicken cells. In vivo inoculation of X-MLV or murine sarcoma virus (MSV) pseudotyped by X-MLV was performed in duck embryos or newborn ducklings (Levy et al., 1982). Evidence of virus persistence has been obtained, but not convincing data concerning their pathogenesis. In mammalian cells, X-MLV-pseudotyped RSV can also replicate in the presence of X-MLV first to a low titer, which increases with passages. Furthermore, an envelope component of ALV-C phenotypically mixed with X-MLV was detected after passaging in mammalian cells.

It is not known how X-MLV contributes to RSV replication in mammalian cells. Obviously, it provides at least a part of the Env glycoprotein required for penetration to mammalian cells. However, additional factors complementing the non-permissiveness of mammalian cells to RSV should be in play. Due to its ability to multiply in such cells, X-MLV can increase and facilitate some post-transcriptional steps such as viral RNA proper splicing and export from the nucleus and/or further steps involving cleavage of viral protein precursors and virion assembly. These facilitating effects of X-MLV should be synchronized with RSV infection, because superinfection of an already RSV-transformed mammalian cell, containing functional proviruses rescuable by fusion with chicken fibroblasts, does not result in virus production (Levy, 1977).

Because X-MLV contribution to RSV replication in mammalian cells increases with passages, a possibility of genetic exchange between these viruses should be taken into account. In every case, these questions should be reinvestigated using presently available efficient tools of molecular biology.

3. Retrovirus heterotransmission among species within the same class

One of the first thoroughly analyzed successful heterotransmissions was achieved by Duran-Reynals (1942), who showed that RSV produces both early and late appearing tumors when inoculated in young ducks. Since then, many

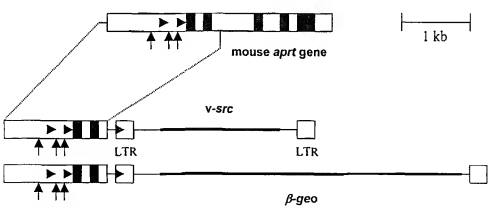


Fig. 2. Cloning of the *aprt* gene CpG island and RSV-based reporter proviruses. Filled boxes represent exons of the *aprt* gene. Vertical arrows denote the position of three Sp1 sites in the CpG island. Filled arrowheads denote transcription starts of the *aprt* gene and 5′ LTR.

Table 1
Trans-species retrovirus transmission

Virus	Species of origin	Original pathogenicity	Transmitted to species	New species pathogenicity	References
SIV_{cpz}	Chimpanzees	Zero to very low	Human	AIDS (HIV-1)	Gao et al., 1999; Weiss and Wrangham, 1999.
SIV_{sm}	Sooty mangabey	Zero to very low	Human	AIDS (HIV-2)	Hirsch et al., 1989; Gao et al., 1992; Sharp et al., 1995; Chen et al., 1996.
BLV	Cattle	Lymphocytosis, leukosis, B-tropic	Rabbit	Wasting disease, immunodeficiency	Burny et al., 1985; Altanerova et al., 1989; Wyatt et al., 1989; Kucerova et al., 1999.
Friend MLV	Mice	Erythroleukemia	Rat	Bone marrow suppression, $Thy1^+$ cell reduction	Mazgareanu et al., 1998.
ALV-C	Chicken	Anemia, probably immunodeficiency	Duck	Wasting disease, immunodeficiency, anemia	Karakoz et al., 1980;
ALV-B ALV-D		Anemia			Smith and Schmidt, 1982.

other retroviruses, including these of mammalian origin, were experimentally transmitted among different mammalian species. These transmissions were monitored mainly by virus oncogenic activity and virus persistence. Because of generally low retrovirus replication in foreign species, additional pathogenic virus activity, such as immunosuppression, usually do not appear. In many cases heterotransmission produced the same symptoms as in the species of the virus origin, but there are well-documented cases of changed virus pathogenicity. Further, we shall deal mainly with the symptoms of wasting disease accompanied by immunodeficiency resulting in increased susceptibility to various infectious agents.

The simian immunodeficiency viruses (SIV) highlight this situation. Generally, in the monkey species, in which SIV is indigenous, it replicates efficiently but does not produce any pathogenic changes. However, when transmitted to some other monkey species, SIV produces the immunodeficiency syndrome. Of special interest is SIV heterotransmission to humans. As is summarized in Table 1, good evidence based on molecular biology and epidemiology has been provided documenting that both human immunodeficiency virus 1 (HIV-1) and HIV-2 represent a consequence of respective SIV_{sm} (sooty mangabey SIV) and SIV_{cpz} (chimpanzee SIV) heterotransmission to humans. There is no doubt about the significance of this finding, which should be taken as a warning against potential danger of retrovirus transgression of species barriers associated with fulminating pathogenic changes. In order to understand these events, comparative data obtained with other members of the retrovirus family should be evaluated.

Interesting observations were made in the case of bovine leukemia virus (BLV) (Table 1). This virus responsible for cattle leukosis when transferred to newborn rabbits triggers clear symptoms of immunodeficiency. Because such a response was not found in other infected species, these observations indicate that rabbits respond to BLV inherently, in an unusual way.

In murine leukemia viruses, variants capable to produce immunodeficiency preferably have been isolated. This includes both Moloney MLV (Saha et al., 1994) and Friend leukemia complex (Faxvaag et al., 1993). In addition, as given in Table 1, Friend MLV transmission to newborn rats resulted in altered pathogenicity characterized by suppression of bone marrow cells, manifesting itself as reduced numbers of $Thy1^+$ cells.

Avian leukosis viruses have not been thoroughly investigated from the point of view of their immunopathogenicity in foreign avian species. As given in Table 1, ALV subgroup C were studied using intraembryonic inoculation both in chicken and ducks. According to data obtained so far, this subgroup produces symptoms of anemia in both species. However, in heterologous duck hosts, a fatal wasting disease together with conspicuous atrophy of the thymus tissue starting the first week after hatching was observed both by a decrease in the relative thymus to body weight, histologically characterized by clear depletion of the thymus cortical layer (Fig. 3) (Stepanets et al., 2000). Microscopically, bursa Fabricii, which constitutes a special B-cell producing organ in birds, also displayed the cortical layer depletion. In agreement with this observation, production of humoral antibodies against *Brucella abortus* antigens was significantly decreased in infected animals, which confirms that immunodeficiency is involved. The nature of this immunodeficiency is being investigated and the character of specific T- and other cell alteration in thymus is monitored by specific antibodies. In addition, changes in lymphoid organs in young chickens intraembryonally inoculated with different ALV subgroups should be investigated. According to our preliminary data, ALV-C does produce symptoms of thymus involution even in chickens. Thus ALV-C provides a suitable comparative system for establishing ways leading to the immunodeficiency syndrome both in homologous and heterologous hosts.

There is not a simple answer to the question why retroviruses in some species behave pathogenically or produce

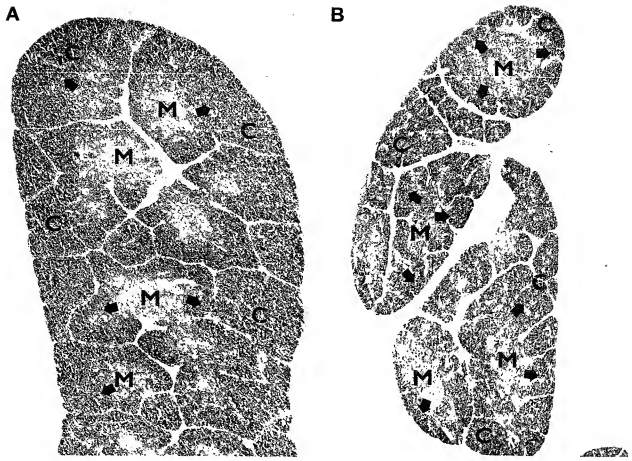


Fig. 3. Thymus sections of 7-day-old ducklings. (A) Thymus of a control animal, mock-injected with tissue culture medium. (B) Thymus of an animal infected in mid-embryogenesis with 10^5 ALV-C infectious virus. Thymic cortical layer is visibly depleted. C, cortex; M, medulla; arrowheads mark the interim between cortex and medulla. Stained with hematoxylin-eosin. Magnification 41×.

new or accentuated pathogenic symptoms. In some cases, such as Friend MLV, selection of virus mutants might be involved, but this seems not to be a general situation, because altered pathogenicity appears very soon after virus inoculation and is not correlated with certain specific virus gene alteration. Therefore other, especially host-specific factors should be taken into account. It is not known so far which of them play a decisive role. Retrovirus toxicity is in the first step related to the outcome of the viral envelope (product of the *env* gene) interaction with target cells, immune and bone-marrow cells included. Therefore, if an infected cell harbors unusually expressed receptors and co-receptors as well as factors required for further steps of virus penetration, they may undergo deterioration or apoptosis. As was discussed in relation to HIV (Fauci, 1996), the degree of cytokine activity and the outcome of interrelation among different cytokine pathways contribute to virus expression and its pathogenicity.

In spite of the fact that heterotransmission of SIV took place probably several times, we are lacking the exact information about this process. In contrast to the experimental systems, natural conditions are probably different, because only low amounts of virus or better virus-infected cells from a donor might have mediated heterotransmission. Such a situation has not been modeled in detail, but it can be predicted that under such conditions the virus becomes efficiently expressed only rarely. Therefore, there might be a bridge connecting trans-class and trans-species retrovirus heterotransmission.

Some experimental data obtained from long-term follow-up of retrovirus persistence (Trejbalová et al., 1999) suggest that provirus silencing discussed in the first and also next chapter might play a role in the control of low-dose infection and long-term retrovirus persistence.

4. Provirus silencing and expression in differentiated cells

In order to investigate the role of cell differentiation in relation to retrovirus infection, several models have been designed, but most thoroughly were analyzed EC cells, which differentiate in vitro. The entrée to this problem was given by the group of Jaenisch (Stewart et al., 1982).

who discovered that MLV integrated in EC cells becomes methylated and unexpressed. In a series of papers from different laboratories (reviewed by Challita et al., 1995) it was established that in order to ensure MLV expression in EC, introduction of a transcription factor Sp1 binding site into LTR is required together with inactivation of the negative control region (NCR) in LTR as well as of the region of the primer binding site. As the third negatively acting element, one out of two direct repeats in LTR was recognized (Hawley et al., 1994). There is still room left for MLV improved expression, as exemplified by insertion into LTR of an antimethylation fragment from the region upstream of the Thy1 gene (Challita et al., 1995).

All the above-mentioned modifications act synergistically. Recently, MLV LTR lacking known negative elements due to deletion spanning most of the 5′ end LTR portion has been constructed (Osborne et al., 1999). As a result of such a deletion, about half of 13 CpG sites within LTR were also lost. Such truncated LTR, essentially stripped of enhancer elements, was employed for generation of a retroviral vector harboring the neomycin resistance (*neo*) reporter gene equipped with internal β-globin promoter, which after infection of EC cells displayed expression in 70% of cells, the highest efficiency obtained so far.

We have not yet reached the end of the journey to optimal retroviral vector function in differentiated cells. It is still possible that additional, so far undefined sequences should be inactivated, altered or inserted. Of special importance might be CpG dinucleotides present in LTR, especially at the start of transcription. Elements acting as antimethylation signals, interfering with silencers or ensuring position-independent gene expression, such as locus control regions (LCR), should be tested for their ability to ensure retrovirus or retroviral vector expression in differentiated cells. Recently, the chicken hypersensitive site 4 (cHS4) of the chicken globin LCR, acting as an insulator, when cloned into MLV LTR was shown to increase the probability of integrated proviruses expression and to decrease the level of de novo methylation of the 5′LTR in murine erythroleukemia (MEL) cells (Rivella et al., 2000). In addition, the human β interferon scaffold attachment region (IFN-SAR), when inserted in retroviral LTR, prevented its methylation and ensured vector expression in a stably transfected line of human T cells. The vector expression has been kept for several months and included also multiple proviral copies (Agarwal et al., 1998; Dang et al., 2000). We can therefore stress the point that the problem of permissiveness of differentiated cells to retrovirus infection goes far beyond EC cells and that other differentiated cells, such as hematopoietic or hepatic cells, should probably require not only prevention of vector downregulation, but even some more specific cell changes like activation of steps triggering the cell cycle (rev. Emerman, 2000).

Dealing with provirus silencing in differentiated cells we focused on provirus methylation as an epigenetic DNA modification described repeatedly in conjunction with provirus downregulation. However, it is not clear so far whether provirus methylation acts as the primary cause or whether it only conserves transcriptional repression. In provirus methylation studies, usually a general increase in CpG methylation has been measured. However, some more precise data have been obtained showing that methylation of one particular CpG has a decisive effect. This is the case of HIV, where methylation of only one CpG at position −143 (in the vicinity of NF-κB and Sp1 binding sites) in HIV LTR results in 70% inhibition of the reporter expression (Bednarik et al., 1990; Schulze-Forster et al., 1990). Similarly, single CpG site methylation in the RSV LTR U5 region nearby the transcription start leads to a clear decrease in provirus expression (Hejnar et al., 1999). As discussed later, the density of CpGs is also of importance.

It has been recognized on other gene models that methylation provides a signal for association with methyl-CpG-binding protein 2 (MeCP2), which, through the adaptor protein Sin3A, recruits the histone deacetylase (HDAC) (rev. Razin, 1998; Ng and Bird, 1999; Knoepfler and Eisenman, 1999). A similar situation was disclosed in the case of MeCP1, which produces a complex composed of the MBD protein containing the methyl-CpG-binding domain and two members of the HDAC family (rev. Bird and Wolffe, 1999). Thus, there is a proven link between CpG methylation and chromatin deacetylation.

How are these findings related to provirus silencing? It should be noted that MeCP1 has been already shown to bind to methylated LTR of myeloproliferative sarcoma virus, suppressing its LTR activity (Boyes and Bird, 1991). This problem has been recently approached by Lorincz et al. (2000). They employed the MLV LTR-driven green fluorescence protein (GFP) gene and followed GFP expression after infection of MEL cells with this vector. Cell clones that displayed silenced GFP were isolated and it was disclosed that early after silencing the proviruses became methylated to a low degree. Such clones could have been reactivated by trichostatin (TSA), which inhibits HDAC-MeCP2 complexes. During prolonged cell cultivation the vector has been increasingly methylated and in the hypermethylated state provirus expression was induced only by combination of TSA and 5-azacytidine (5-azaC), which acts as a demethylation agent. These results suggest that provirus downregulation is a dynamic process and that the possibility of its reactivation depends upon the density of methylation.

All the data obtained so far point to the important role of methylation in provirus silencing in general, and therefore strategies preventing methylation of retroviral vectors as well as putative blocking of methyl-CpG-binging proteins should contribute to more efficient gene therapy applied to differentiating or differentiated cells.

There is, in addition, an important question: whether and why retroviruses are more efficiently recognized and silenced than any foreign DNA introduced in a genome. In other words, does there exist a cell genome surveillance mechanism (analogous but not homologous to immunity)

that ensures downregulation of retroviral sequences? Methylation has been already proposed to fulfil such a duty (Doerfler, 1991; Yoder et al., 1997). However, some signals should be involved that attract methyltransferase to an integrated provirus, or even to a specific DNA structure common to viral integration intermediates (Bestor, 1987). Such signals might be provided by flanking chromosomal sequences or by the proviral structure itself. Especially LTR could be recognized as unusual direct repeats. Similar structures in lower eukaryotes trigger gene silencing (rev. Wolffe and Matzke, 1999). Furthermore, MLV LTR binding zinc finger transcription factor YY-1 (Flanagan et al., 1992) was shown to represent a homolog of Sin3 (Yang et al., 1996), which has been already identified as a protein complexing with HDAC involved in formation of transcriptionally inactive chromatine. It should be also taken into account that in some heterologous and differentiated cells, proteins activating LTR might be underrepresented or that such cells produce altered isomorphic proteins, which could be inactive and/or could interfere with factors required for LTR activation. Therefore, retroviral genomes should be screened also from the point of view of sequences and DNA-protein complexes which might contribute to provirus silencing, of course, in context with flanking chromosomal DNA. New techniques such as inverse polymerase chain reaction should facilitate this demanding task.

Acknowledgements

The authors highly acknowledge the excellent editorial help of Šárka Takáčová. These studies, which involve several interconnected areas of research, were supported by Grant No. 1R03 TW00 155-01A1 (Fogarty International Research Collaboration Award), Grants No. 301/94/0713, No. 203/96/0883, No. 312/96/K205, No. 312/98/0825 and No. 312/97/P082 from the Grant Agency of the Czech Republic.

References

Agarwal, M., Austin, T.W., Morel, F., Chen, J., Böhnlein, E., Plavec, I., 1998. Scaffold attachment region-mediated enhancement of retroviral vector expression in primary T cells. J. Virol. 72, 3720–3728.

Altanerova, V., Ban, J., Altaner, C., 1989. Induction of immune deficiency syndrome in rabbits by bovine leukaemia virus. AIDS 3, 755–758.

Anderssen, S., Sjettem, E., Svineng, G., Johansen, T., 1997. Comparative analysis of LTR of the ERV-H family of primate-specific retrovirus-like elements isolated from marmoset African green monkey. and man. Virology 234, 14–30.

Bednarik, D.P., Cook, J.A., Pitha, P.M., 1990. Inactivation of the HIV LTR by DNA CpG methylation Evidence for a role in latency. EMBO J. 9, 1157–1164.

Bestor, T.H., 1987. Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. Nucleic Acids Res. 15, 3835–3843.

Bird, A.P., Wolffe, A.P., 1999. Methylation-induced repression - belts, braces, and chromatin. Cell 99, 451–454.

Boyes, J., Bird, A., 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 64, 1123–1134.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., Cedar, H., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature 371, 435–438.

Burny, A., Bruck, C., Cleuter, Y., Couez, D., Deschamps, J., Ghysdael, J., Gregoire, D., Kettmann, R., Mammerickx, M., Marbaix, G., 1985. Bovine leukaemia virus, a versatile agent with various pathogenic effects in various animal species. Cancer Res. 45, 4578s–4582s.

Challita, P.-M., Skelton, D., El-Khoueiry, A., Yu, X.-J., Weinberg, K., Kohn, D.B., 1995. Multiple modifications in *cis* elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinome cells. J. Virol. 69, 748–755.

Chen, Z., Telfer, P., Gettie, A., Reed, P., Zhang, L., Ho, D.D., Marx, P.A., 1996. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. J. Virol. 70, 3617–3627.

Dang, Q., Auten, J., Plavec, I., 2000. Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. J. Virol. 74, 2671–2678.

Doerfler, W., 1991. Patterns of DNA methylation - evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. Biol. Chem. Hoppe-Seyler 372, 557–564.

Duran-Reynals, F., 1942. The reciprocal infection of ducks and chickens with tumor inducing viruses. Cancer Res. 2, 343–369.

Emerman, M., 2000. Learning from lentiviruses. Nature Genet. 24, 8–9.

Fauci, A.S., 1996. Host factors and the pathogenesis of HIV-induced disease. Nature 384, 529–534.

Faxvaag, A., Dai, H.Y., Aarseth, H., Dalen, A.B., 1993. A low oncogenic variant of friend murine leukemia virus with strong immunosuppressive properties. Arch. Virol. 131, 265–275.

Fincham, V.J., Wyke, J.A., 1991. Differences between cellular integration sites of transcribed and nontranscribed Rous sarcoma proviruses. J. Virol. 65, 461–463.

Flanagan, J.R., Becker, K.G., Ennist, D.L., Gleason, S.L., Driggers, P.H., Levi, B.Z., Appella, E., Ozato, K., 1992. Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. Mol. Cell. Biol. 12, 38–44.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Hahn, B.H., 1992. Human infection by genetically diverse SIV_{SM}-related HIV-2 in West Africa. Nature 358, 495–499.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.M., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee. *Pan troglodytes troglodytes*. Nature 397, 436–441.

Hawley, R.G., Lieu, F.H., Fong, A.Z., Hawley, T.S., 1994. Versatile retroviral vectors for potential use in gene therapy. Gene Ther. 1, 136–138.

Hejnar, J., Plachý, J., Geryk, J., Machoň, O., Trejbalová, K., Guntaka, R.V., Svoboda, J., 1999. Inhibition of the Rous sarcoma virus long terminal repeat-driven transcription by in vitro methylation: different sensitivity in permissive chicken cells versus mammalian cells. Virology 255, 171–181.

Hirsch, V.M., Olmsted, R.A., Murphy-Corb, M., Purcell, R.H., Johnson, P.R., 1989. An African primate lentivirus (SIV_{sm}) closely related to HIV-2. Nature 339, 389–392.

Karakoz, I., Geryk, J., Svoboda, J., 1980. In vivo effect of three transformation-defective mutants of subgroup C avian sarcoma viruses. In: Zatz-zili, S., Sponar, J. (Eds.), DNA – Recombination, Interactions and Repair. Pergamon Press, Oxford and New York, pp. 435–440.

Knoepfler, P.S., Eisenman, R.N., 1999. Sin meets NuRD and other tails of repression. Cell 99, 447–450.

Kuzerova, L., Altanerova, V., Altaner, C., Boris-Lawrie, K., 1999. Bovine

leukemia virus structural gene vectors are immunogenic and lack pathogenicity in a rabbit model. J. Virol. 73, 8160–8166.
Levy, J.A., 1977. Murine xenotropic type C viruses III. Phenotypic mixing with avian leukosis and sarcoma viruses. Virology 77, 811–825.
Levy, J.A., 1978. Xenotropic type C viruses. Curr. Top. Microbiol. Immunol. 79, 111–213.
Levy, J.A., Oleszko, O., Dimpfl, J., Lau, D., Rigdon, R.H., Jones, J., Avery, R., 1982. Murine xenotropic type C viruses IV. Replication and pathogenesis in ducks. J. Gen. Virol. 61, 65–74.
Lorincz, M.C., Schübeler, D., Goeke, S.C., Walters, M., Groudine, M., Martin, D.I.K., 2000. Dynamic analysis of proviral induction and de novo methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. Mol. Cell. Biol. 20, 842–850.
Machoň, O., Strmen, V., Hejnar, J., Geryk, J., Svoboda, J., 1998. Sp1 binding sites inserted into the Rous sarcoma virus long terminal repeat enhance LTR-driven gene expression. Gene 208, 73–82.
Macleod, D., Charlton, J., Mullins, J., Bird, A.P., 1994. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282–2292.
Mazgareanu, S., Muller, J.G., Czub, S., Schimmer, S., Bredt, M., Czub, M., 1998. Suppression of rat bone marrow cells by Friend murine leukemia virus envelope proteins. Virology 242, 357–365.
Mummaneni, P., Bishop, P.L., Turker, M.S., 1993. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. J. Biol. Chem. 268, 552–558.
Nelson, D.T., Goodchild, N.L., Mager, D.L., 1996. Gain of Sp1 sites and loss of repressor sequences associated with a young, transcriptionally active subset of HERV-H endogenous long terminal repeats. Virology 220, 2213–2218.
Ng, H.-H., Bird, A., 1999. DNA methylation and chromatin modification. Curr. Opin. Genet. Dev. 9, 158–163.
Osborne, C.S., Pasceri, P., Singal, R., Sukonnik, T., Ginder, G.D., Ellis, J., 1999. Amelioration of retroviral vector silencing in locus control region β-globin-transgenic mice and transduced F9 embryonic cells. J. Virol. 73, 5490–5496.
Prince, V.E., Rigby, P.W.J., 1991. Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. J. Virol. 65, 1803–1811.
Razin, A., 1998. CpG methylation, chromatin structure and gene silencing – a three-way connection. EMBO J. 17, 4905–4908.
Rivella, S., Callegari, J.A., May, C., Tan, C.W., Sadelain, M., 2000. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. J. Virol. 74, 4679–4687.
Ryndich, A., Kadi, R., Geryk, J., Zoubak, S., Svoboda, J., Bernardi, G., 1991. The isopycnic compartmentalized integration of Rous sacoma virus sequences. Gene 106, 165–172.
Saha, K., Yuen, P.H., Wong, P.K.Y., 1994. Murine retrovirus-induced depletion of T cells is mediated through activation-induced death by apoptosis. J. Virol. 68, 2735–2740.
Schulze-Forster, K., Götz, F., Wagner, H., Kroger, H., Simon, D., 1990. Transcription of HIV 1 is inhibited by DNA methylation. Biochem. Biophys. Res. Commun. 168, 141–147.
Sharp, P.M., Robertson, D.L., Hahn, B.H., 1995. Cross-species transmission and recombination of 'AIDS' viruses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 349, 41–47.
Sjottem, E., Anderssen, S., Johansen, T., 1996. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. J. Virol. 70, 188–198.
Smith, R.E., Schmidt, E.V., 1982. Induction of anemia by avian leukosis viruses of five subgroups. Virology 117, 516–518.
Stepanets, V., Vernerová, Z., Vilhelmová, M., Geryk, J., Hejnar, J., Weichold, F.F., Svoboda, J., 2000. Intraembryonic avian leukosis virus subgroup C (ALV-C) inoculation producing wasting disease in ducks soon after hatching (submitted).
Stewart, C.L., Stuhlmann, H., Jahner, D., Jaenisch, R., 1982. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 79, 4098–4102.
Svoboda, J., 1986. Rous sarcoma virus. Intervirology 26, 1–60.
Svoboda, J., 1998. Molecular biology of cell non-permissiveness to retroviruses. Has the time come? Gene 206, 153–163.
Trejbalová, K., Gebhard, K., Vernerová, Z., Dušek, L., Geryk, J., Hejnar, J., Haase, A.T., Svoboda, J., 1999. Proviral load and expression of avian leukosis viruses subgroup C in long-term persistently infected heterologous hosts (ducks). Arch. Virol. 144, 1779–1807.
Weiss, R.A. Wrangham, R.W., 1999. From *pan* to pandemic. Nature 385–386.
Wolffe, A.P., Matzke, M.A., 1999. Epigenetics: regulation through repression. Science 286, 481–486.
Wyatt, C.R., Wingett, D., White, J.S., Buck, C.D., Knowles, D., Reeves, R., Magnuson, N.S., 1989. Persistent infection of rabbits with bovine leukemia virus associated with development of immune dysfunction. J. Virol., 63, 4498–4506.
Yang, W.M., Inouye, C., Zeng, Y., Bearss, D., Seto, E., 1996. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc. Natl. Acad. Sci. USA 93, 12845–12850.
Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13, 335–340.
Young, W.-B., Lindberg, G.L., Link Jr, C.J., 2000. DNA methylation of helper virus increases genetic instability of retroviral producer cells. J. Virol. 74, 3177–3187.

TAB 18

This material may be protected by Copyright law (Title 17 U.S. Code)

ELSEVIER

Mutation Research 386 (1997) 119–130

Reviews in Mutation Research

Formation of methylation patterns in the mammalian genome

Mitchell S. Turker [a,*,1], Timothy H. Bestor [b]

[a] *Department of Pathology, Markey Cancer Center, University of Kentucky, Lexington, KY 40536, USA*
[b] *Department of Genetics and Development, Columbia University, New York, NY 10032, USA*

Abstract

Cytosine methylation in mammals is an epigenetic modification required for viability of the developing embryo. It has been suggested that DNA methylation plays important roles in X-chromosome inactivation, imprinting, protection of the genome from invasive DNA sequences, and compartmentalization of the genome into active and condensed regions. Despite the significance of DNA methylation in mammalian cells, the mechanisms used to establish methylation patterns during development are not understood. This review will summarize the current state of knowledge about potential roles for *cis*- and *trans*-acting factors in the formation of methylation patterns in the mammalian genome.

Keywords: Mamalian genome; Methylation pattern

1. Changes in methylation patterns during mammalian development

The mammalian genome is heavily methylated at cytosine residues within the dinucleotide CpG. During early mammalian development gamete-specific methylation patterns are converted to somatic cell-specific methylation patterns via poorly-understood waves of demethylation and de novo methylation [1]. Methylation levels are reduced dramatically between the 4 cell stage and the 64 cell stage, then restored to high levels via de novo methylation commencing at the time of implantation. For a given locus, tissue-specific methylation patterns may undergo further changes during development. The mechanisms used by mammalian cells to establish methylation patterns are not known. The subject is of special interest as aberrant DNA methylation patterns have been linked to altered gene expression in certain genetic diseases [2–4] and cancers [4,5].

The distribution of methylated cytosines in the mammalian genome is not random. Promoter regions of actively transcribed genes, which are frequently rich in the CpG dinucleotide sequence, are almost invariably unmethylated [6,7]. Some genes, such as those subject to genomic imprinting [8] and genes on the X-chromosome [9], can also exist in a repressed state characterized by heavily methylated promoter regions. The coding region of many genes, regardless of expression status, are moderately or heavily methylated. Intergenic regions, which consist predominately of moderately and highly repeated sequence elements, are also heavily methylated in most tissues. An example of a well-characterized repetitive sequence is the primate Alu element [10]. This

* Corresponding author. Tel.: +1 (503) 494-4273; Fax: +1 (503) 494-4278.
[1] CROET, L606, Oregon Health Sciences University, 3181 S. W. Sam Jackson Park Road, Portland, OR 97201, USA.

1383-5742/97/$17.00 Copyright © 1997 Elsevier Science B.V. All rights reserved.
PII S1383-5742(96)00048-6

element is relatively CpG-rich and heavily methylated in all somatic tissues [11,12] and female gametes [13]. Methylation is linked to transcriptional silencing of Alu elements [14].

While much is uncertain, there are two properties of methylation patterns that can be regarded as well-proven. First, methylation of CpG sites within mammalian promoters represses transcription, often to undetectable levels. Second, methylation patterns are subject to clonal inheritance in most cases. These properties suit methylation patterns for a large number of biological roles, although none of the hypothetical functions can be regarded as proven or disproven.

2. Why is the mammalian genome methylated?

Mice with targeted mutations at the gene (*Dnmt*) that encodes the one known form of DNA methyltransferase die as embryos [15,16], demonstrating that disruption of methylation patterns is incompatible with normal mammalian development. However, the nature of the essential function could not be deduced from the phenotype of the mutant embryos.

Methylation patterns have been proposed to regulate gene expression during development [17,18], to control DNA replication [19], and to play an important role in DNA repair [20]. Another hypothetical function of methylation patterns is genome defense, and support for this hypothesis is accumulating. It was proposed that DNA methylation in mammals serves to defend the genome from the proliferation of parasitic sequence elements [21]. Initially, this mechanism may have served to inactivate exogenous DNA sequences such as viral genomes via methylation-associated gene silencing [21,22]. At least a remnant of this mechanism exists as evidenced by the observation of retroviral inactivation in cultured embryonic cells and in the developing embryo [23]. As the mammalian genome expanded, this protective function could have been further adapted to block expression of the numerous repeated sequences and retroposons that comprise the bulk of the mammalian genome [21]. As noted above, such sequences are heavily methylated in mammalian cells. Eventually, the protective function could have been extended to single copy loci as a mechanism to control gene expression in a tissue-specific manner [7,21] or even to control expression of imprinted genes [24]. Moreover, compartmentalization of the large mammalian genome into regions that are methylated and condensed and regions that are unmethylated and in an open chromatin conformation could facilitate the ability of transcription factors to scan the genome in their search for appropriate targets [21].

3. The mammalian DNA methyltransferase

At the present time only a single mammalian form of DNA MTase has been isolated from mammals; a very similar enzyme has been identified in amphibians, echinoderms, and plants. This enzyme is found in essentially all dividing cell types. The mammalian DNA MTase is divided into two major domains which are linked by a stretch of alternating glycyl and lysyl residues (Fig. 1). The 500 amino acid carboxy terminus constitutes the catalytic domain. This domain has extensive similarity with bacterial m5C DNA MTases which are known to act by 'flipping out' the target cytosine from the DNA helix to perform the methylation reaction [25]. The 1000 amino acid amino-terminal portion of mammalian DNA MTase has multiple regulatory functions. Significantly, removal of the regulatory domain from the catalytic domain results in an enzyme that has an equal capacity to recognize and methylate hemimethylated and unmethylated substrates [26], suggesting that the regulatory domain normally acts to suppress the de novo methylation reaction. This property is likely to be at least partially responsible for the clonal propagation of methylation patterns. The observation of an inhibitory function in the regulatory region further suggests that de novo methylation is controlled by cell-type-specific cellular proteins and/or specific DNA structures, which interact with this region, rather than altered forms of DNA MTase [27].

A Zn^{2+} binding site in the regulatory region (Fig. 1) has been implicated in the ability of the DNA MTase to distinguish methylated and hemimethylated sites. This sequence is shared by mammalian DNA MTase and the mammalian homolog of the

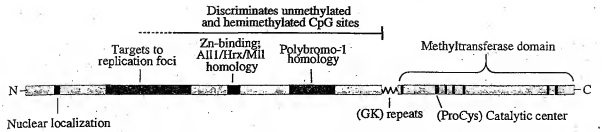


Fig. 1. Sequence motifs and functional domains in mammalian DNA methyltransferase. The protein comprises two domains: a C-terminal catalytic domain that is closely related to bacterial restriction methyltransferases, and a larger N-terminal domain that has several regulatory roles. These include discrimination of unmethylated and hemimethylated CpG sites, nuclear localization, and cell cycle-dependent association with replication foci. Two regions are closely related to vertebrate homologues of proteins that are involved in the clonal transmission of patterns of gene expression in Drosophila; these homologues are ALL1/Hrx/MLL (a trithorax homologue; [28]) and Polybromo-1 (a Brahma homologue; GenBank X90849). These similarities link the two systems known to be involved in the clonal transmission of states of gene expression during metazoan development.

Drosophila trithorax (*trx*) gene. The human homolog has been termed HRX, ALL-1, or MLL because it is frequently found to be involved in translocation rearrangements in acute lymphocytic leukemia [28]. Drosophila trithorax is one of a large group of proteins that are involved in the clonal propagation of states of gene expression, members of the trithorax group maintain the active state, while chromosomal proteins of the polycomb group confer heritable repression [6]. An important target of the trithorax and polycomb proteins are the homeotic genes, which confer cell identity along the anterior-posterior axis. The polycomb group and trithorax group proteins do not establish zones of expression of the homeotic genes but are required for maintenance of zones of expression during development. The function of the polycomb/trithorax proteins may therefore be closely analogous to that of heritable methylation patterns, and methylation patterns may increase the heritability of states of gene expression in higher eukaryotes [6].

It has been shown recently that inactivation of the mouse Hrx/All-1/Mll gene results in altered expression of homeotic (*Hox*) genes [29]. Anterior boundaries of Hoxa-7 and Hoxc-9 were shifted posteriorly in heterozygotes. These mice were viable, although they exhibited phenotypic abnormalities including small size at birth and growth retardation. Hoxa-7 and Hoxc-9 expression were abolished in homozygous deficient animals, which died by embryonic day 10.5. The observation of homology between the Hrx/All1/Mll and DNA MTase proteins in a region used by the DNA MTase to recognize methylation status suggests that the Hrx/All1/Mll protein has evolved to respond differently to unmethylated, hemimethylated, or symmetrically methylated target genes. One test of this possibility will be to determine if the extreme phenotype noted in Mll homozygous deficient animals is partially rescued in a methylation deficient background.

An additional region of the regulatory domain acts to target the enzyme to replication foci, presumably to increase the fidelity of transmission of methylation patterns [30]. Despite the clear ability of the mammalian DNA MTase to localize to the nucleus, it is not found there at early times in development [31]. A very large maternal store of DNA MTase is localized in the cytoplasm from the one cell stage to the four cell stage. As mentioned above, the levels of genomic methylation begin dropping at about the four cell stage. However, at the eight cell stage, when methylation levels are continuing to drop, significant amounts of DNA MTase begin to localize to the nucleus. By the blastocyst stage, most DNA MTase is again found in the cytoplasm where it remains until some time after implantation. In the 5.5 day egg cylinder, DNA MTase has again moved into the nucleus where it will remain in the developing cells of the embryo proper [32]. Significant levels of de novo methylation occur when the blastocyst implants. If it is assumed that the 8 cell embryo lacks factors that allow DNA MTase to act as a de novo

methylation enzyme, we can speculate that maintenance methylation of critical sites is necessary in the 8 cell embryo. According to this speculation, CpG sites that are unmethylated at the 8 cell stage would be unaffected by the presence of DNA MTase. Regardless of the exact reasons, the observation that DNA MTase is shuttled between the nucleus and cytoplasm during early development clearly demonstrates an unusual form of post-transcriptional regulation. A second unusual mechanism that regulates DNA MTase levels has been observed in pachytene spermatocytes. In these cells the commonly observed 5.2 kb form of DNA MTase mRNA is lost and a novel transcript of 6.2 kb appears [33]. The larger transcript is not translated, resulting in the absence of DNA MTase protein in pachytene cells [34].

4. *Cis*-acting elements in the mammalian genome that promote methylation

Silva and White [35] reported that methylation patterns could vary for allelic sites on homologous chromosomes and that the chromosome-specific methylation patterns were heritable. They further reported that methylation patterns in sperm DNA preparations differed from those observed in somatic cells, indicating that the somatic cell-specific methylation patterns were re-established in each generation. In total, their data suggested that chromosome-specific *cis*-acting elements accounted for the formation of methylation patterns at the loci studied.

One prediction from Silva and White's work was that a methylation pattern specified by a *cis*-acting element should be reproducible in different cell populations and/or different individuals. Reports that support this prediction include the observation of a specific methylation pattern in all somatic tissues in the region upstream of the mouse *aprt* gene [36,37] (Fig. 2) and a similar observation for the human tumor necrosis factor α and β genes [38]. In this latter example, tissue-specific differences were noted, but a methylation pattern for the cells of a given tissue was consistent in different individuals, suggesting that other factors can further influence the formation of methylation patterns. It is important to note that some CpG sites exhibit clonally inherited mosaic patterns of methylation [36,39] (Fig. 2), suggesting that, once formed, methylation patterns are not necessarily fixed, but instead may still involve an active process of demethylation and remethylation that allow an equilibrium to be achieved.

The above reports represent indirect evidence that *cis*-acting elements are involved in the formation of methylation patterns in mammalian cells. Direct evidence requires the isolation of these elements, their

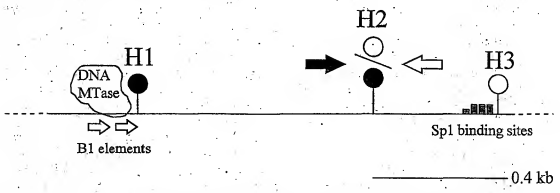


Fig. 2. Methylation pattern in the *aprt* gene region. Methylation levels at three *HpaII/MspI* sites (H1–H3) are represented by bubble figures. A filled bubble indicates a methylated CpG site and an open bubble indicates an unmethylated site. Filled and open arrows surrounding H2 site represent opposing forces of methylation spreading and maintenance of a methylation free region, respectively. Tandem B1 elements are shown as open arrows with the arrowheads indicating 5′ end of each B1 element. Sp1 binding sites are indicated by hatched boxes with a non-consensus site shown as a smaller box. These sites are not drawn to scale. See text for description of how the methylation pattern is formed and maintained.

propagation in a non-methylated state, and demonstration that they can specify de novo methylation when introduced back into the appropriate genome. In a non-mammalian system, such evidence has been obtained by Selker and colleagues [40] in work with the fungus *N. crassa*. Their studies have demonstrated that discrete regions of the *N. crassa* genome can act as portable de novo methylation signals. These signals represent the remnants of duplicated endogenous 5S RNA sequences consistent with the defense hypothesis for DNA methylation. Other *N. crassia* sequences, when duplicated in meiotic cells, can also undergo de novo methylation followed by a bizarre process termed RIP (repeat induced point mutation) in which numerous mutations are introduced into the duplicated sequence. It appears that a critical number of point mutations is required for the duplicated region to act as a portable methylation signal in somatic cells [41]. Therefore, *N. crassa* has at least two mechanisms to de novo methylate a duplicated sequence. The first process, which occurs in meiotic cells, can recognize and methylate any duplicated sequence. The second process, which occurs in somatic cells, can specifically recognize the now mutated sequence as an independent *cis*-acting methylation signal.

Szyf et al. [42] reported that the adrenal-specific gene steroid 21-hydroxylase became de novo methylated when introduced into the cultured adrenocortical cells, but not when introduced into cultured fibroblasts. This result suggested both the presence of a *cis*-acting methylation signal on the genomic fragment tested and cell type specificity for the de novo methylation reaction. Hasse and Schultz [43] demonstrated that fragments from the rat α-fetoprotein control region (i.e., a region including the promoter and enhancer elements) could enhance de novo methylation of a reporter DNA fragment. Some of these fragments contained repetitive DNA elements, suggesting that such elements act as *cis*-acting methylation signals. Although both of the above reports were supportive of the notion that *cis*-acting elements could signal DNA methylation, in neither case was an endogenous methylation pattern recreated by transfection of a DNA fragment. Moreover, both groups used the calcium phosphate method of DNA transfection, which often leads to high copy numbers per transfectant, raising the possibility that copy number could have played a role in the observed levels of de novo methylation.

The first demonstration that an endogenous mammalian methylation pattern could be reproduced by transfection was provided by Turker et al. [44] in an experiment in which an unmethylated (at CpG sites) plasmid containing the mouse *aprt* gene and its upstream region was transfected by electroporation into a cultured mouse embryonal carcinoma cell line. Electroporation was used to keep a low copy number for the integrated plasmids. In essentially all cases, the well defined methylation pattern for the endogenous upstream region (Fig. 2) was reproduced on the integrated plasmids. The formation of the appropriate methylation pattern was independent of the site of plasmid integration, which is consistent with the presence of a *cis*-acting element. This pattern was not formed, however, if the plasmids were transfected into differentiated progeny of the embryonal carcinoma cell line, suggesting the requirement for one of more embryonal-specific factors.

A deletion analysis localized the *cis*-acting methylation signal responsible for the formation of the upstream methylation pattern to a DNA fragment of approximately 0.8 kbp [45]. This element has been termed a 'methylation center'. Recent work suggests that the bulk of the methylation center signal emanates from tandem B1 repetitive elements (Mummaneni and Turker, in preparation). Similarly, it has been proposed by Magawu and Jones [46] that a human Alu element is responsible for a high level of methylation of a frequently mutated CpG site in the seventh exon of the human *p53* gene. Alu and B1 elements are homologous [47]. Alu elements are imperfect dimers of a monomeric sequence whereas the B1 element represents the monomer. Both elements are believed to be derived from reverse transcription of 7SL RNA [48].

CpG sites that are proximal or distal to the *aprt* methylation center are dependent upon its presence for methylation to occur [45]. Therefore, the methylation signal must somehow spread from the methylation center. Work by Toth et al. [49] examining the spreading of methylation from preimposed methylated CpG sites in the adenovirus genome in stably transfected cultured cells demonstrated that the spreading event is gradual requiring a large number of cell generations. This result suggests that the

phenomenon of methylation spreading is not a processive reaction that occurs via a single pass of the DNA MTase.

5. *Cis*-acting elements as imprinting boxes

Allele-specific methylation is a hallmark of imprinted genes [8]. An excellent example of this rule is the mouse *Igf-2r* gene, which is non-expressed and methylated in the promoter region (region 1) when inherited from the father and expressed and methylated in an intronic region (region 2) when inherited from the mother. The maternally inherited promoter region of Igf-2r remains unmethylated. The maternally inherited methylation pattern is established during gametogenesis and it is believed to be dominant to the paternally inherited pattern, which is established during embryogenesis [50]. Therefore, methylation of region 2 precludes methylation of region 1. It has been proposed that repeated elements within region 2 act to attract methylation by virtue of their resemblance to 'foreign DNA' [51]. This proposal assumes that the putative host defense function for DNA methylation has evolved to play a role in the process of imprinting. Consistent with this hypothesis is work by Chaillet et al. [52] with an imprinted mouse transgene termed RSVIgmyc. This transgene expresses a *c-myc* oncogene, derived from a murine immunoglobin heavy chain a region/*c-myc* (IgA/*c-myc*) translocation, only in paternally inherited heart tissue where the transgene in unmethylated. The same transgene is methylated and not expressed in maternally inherited heart tissue. Using a series of deletion constructs in transgenic animals Chaillet et al. [52] were able to localize the imprinting signal to repeat sequences that make up the 3′ portion of the immunoglobin heavy chain a region. It has also been reported that insertion of the defective retroviral IAP (intercisternal A particle) element into the mouse agouti locus will result in imprinting of this locus [53,54]. IAP elements are heavily methylated in somatic cells [55]. The LTR (long terminal repeat) region of the IAP element, in which allele-specific methylation has been noted [54] has repeated motifs [51]. Finally, the presence of a *cis*-acting imprinting center on human chromosome 15 has been inferred from the observation that inherited microdeletions in the region 15q11-13 will result in the Angelman and Prader-Willi syndromes [56]. Both syndromes are due to inappropriate expression of one or more imprinted genes in this region. It should be noted that the existence of imprinting boxes cannot be regarded as proven, since no sequence from the vicinity of an endogenous imprinted gene can fully reproduce imprinting behavior when present as a transgene.

6. What is the *cis*-acting signal?

The above sections are supportive of the hypothesis that *cis*-acting elements can signal DNA methylation and suggest that repetitive regions may acts as signals. If it is assumed that the *cis*-acting elements act to attract the DNA MTase, it becomes important to ask what the DNA MTase recognizes. Based on work with supercoiled plasmid DNA, Bestor [57] suggested that alternative secondary structures could attract or repel mammalian DNA MTase. Recent work by Christman et al. [58] has also suggested that secondary structure of single stranded DNA may act in *cis* to signal DNA methylation. It has been further shown that unusual DNA structures, particularly those that disrupt and weaken the strength of cytosine base-pairing or stacking, will act as efficient in vitro signals for de novo methylation by a preparation of human DNA MTase [59]. This observation may reflect the necessity of the DNA MTase to flip the cytosine out from the double helix for the methylation reaction to occur.

Recently, investigators have turned to the gene responsible for the fragile X-syndrome, FMR-1 to determine how *cis*-acting sequences can attract the DNA MTase. The molecular defect underlying the fragile X-syndrome is expansion of a trinucleotide repeat, CGG, from a copy number of 5–52 to a copy number of greater than 200. Expansion of this region is almost always linked to its de novo methylation [60] and transcriptional inactivation of FMR-1 [61]. Therefore, the expanded trinucleotide repeat region of FMR-1 represents an example in which a *cis*-acting methylation signal is created from a region that does not normally act as such a signal. Several reports [62–64] have shown that the trinucleotide

repeat region forms hairpin structures that can attract de novo methylation by the human DNA MTase [62]. It has been proposed that the C-rich hairpin is specifically methylated by the enzyme because the cytosine residues in the hairpin will have considerable freedom to flip in and out of the structure [62].

Based on the above data, it is tempting to speculate that repetitive elements which act to attract DNA methylation may also form secondary structures. Although there is no evidence at the DNA level to support this possibility, transcripts from repetitive elements such as B1 [65] and Alu [48] can readily form complex secondary structures. Alternatively, RNA directed de novo DNA methylation has been demonstrated for a viroid element in a plant system [66] and it has been shown that the XIST (X inactivation-specific transcript) RNA acts in *cis* to inactivate one of the two X-chromosomes in female cells [67]. Therefore, secondary structures at either the DNA or RNA levels might play roles in signaling DNA methylation. Finally, it is possible that a Holliday junction intermediate can promote de novo methylation of an unmethylated sequence annealed with a methylated homolog. This possibility is covered in detail elsewhere [4].

7. *Cis*-acting elements in the mammalian genome that block de novo methylation

Promoter regions, particularly if they rich in the CpG dinucleotide (i.e., CpG islands), are inactivated when heavily methylated. If methylation can truly spread from methylation center regions, CpG-rich promoters near such centers should be at risk for methylation associated inactivation. It has therefore been suggested that these CpG rich regions should be protected from de novo methylation [68]. Work by Szyf et al. [69] with the Thy-1 promoter region and Hasse et al. [70] with the metallothionein promoter region suggested that *cis*-acting elements could indeed block the process of de novo methylation in cultured embryonal cells. CpG island sequences have also been shown to be resistant to methylation in vitro by purified mammalian DNA MTase [71].

The presence of a methylation center upstream of the mouse *aprt* promoter [45], coupled with the observation that the promoter was sensitive to methylation-associated gene inactivation [72] suggested that a mechanism existed that blocked the spread of methylation into the *aprt* promoter [37]. This promoter is comprised of 3 consensus and one non-consensus Sp1 binding sites (Fig. 2). Only the two 3′ Sp1 binding sites are required for maximal gene expression [73]. In experiments designed to determine if the methylation center could inactivate the *aprt* promoter, it was found that removal of the two 5′ Sp1 binding sites (Fig. 2) was necessary for the methylation-associated inactivation event to occur [74]. This result suggested that the two 5′ Sp1 binding sites were necessary to block the spreading of de novo methylation into the two 3′ Sp1 elements, thereby preventing their inactivation. Moreover, it also suggested that the *aprt* promoter was duplicated during evolution to prevent high frequency inactivation from occurring [37]. Independent work by Macleod et al. [75] with the mouse *aprt* promoter and Brandeis et al. [76] with the hamster *aprt* promoter demonstrated that the presence of Sp1 binding sites was indeed necessary to block the spreading of methylation in embryonal stem cells and in transgenic embryos. Mutation of these sites to eliminate Sp1 binding also eliminated their ability to block the methylation event, suggesting that the presence of the Sp1 transcription factor provided a physical block to methylation spreading.

It is possible that the presence of occupied Sp1 binding sites will influence nearby CpG sites to remain unmethylated. If so, CpG sites caught between the *aprt* methylation center and the *aprt* promoter will be subject to the opposing forces of methylation and demethylation which may explain how partial methylation patterns [36] are established and maintained (Fig. 2).

8. Modifying factors

As suggested above, the Sp1 protein may act to prevent methylation of promoter regions. Other factors may exist that do not block methylation, but instead actively promote loss of methylated residues during development. This possibility has been suggested by work demonstrating demethylating activities in the extracts of differentiating mouse myoblasts and developing chick embryos [77]. The

demethylation event occurs by the replacement of 5-methylcytosine with cytosine [78]. Interestingly, a 5-methylcytosine-DNA glycosylase has been recently isolated from 12 day old chick embryos [79]. In contrast to the ubiquitous Sp1 transcription protein, it is likely that the demethylating enzymes will be found in only some cells types. Active demethylation of methylated constructs has been observed in embryonal stem cells and in developing embryos [80], though interpretation of these experiments is complicated by the presence of Sp1 binding sites on the tested constructs. Finally, recent experiments have suggested that RNA is involved in the demethylation reaction [80a].

A second line of evidence for cell-type-specific modifying factors is that methylation patterns can vary in different tissues for a given locus or related loci. As mentioned above, most copies of the human Alu element are heavily methylated in somatic tissues. These elements are also heavily methylated in female germ cells [13]. In contrast, most Alu elements are not methylated in mature sperm cells, suggesting the presence of a modifying factor in these cells that blocks Alu element methylation. Recently an Alu binding protein was isolated from human sperm (termed SABP for sperm Alu binding protein) and shown to block the ability of the prokaryotic *Sss*I DNA MTase (which methylates all CpG sites) to methylate Alu sequences in vitro [81]. Further work will be required to determine if this protein can act similarly in vivo and to determine if it is absent in somatic cells and female germ cells. The presence of germ cell-specific modifying factors that recognize specific *cis*-acting sequences is a requirement of most imprinting models, particularly those that predict a role for 'foreign-like' DNA sequences as the *cis*-acting imprinting boxes.

Genetic experiments have also provided very strong evidence supporting the existence of modifying factors. Allen et al. [82] reported that transgene expression and methylation levels varied in different strain backgrounds. For example, methylation levels were highest in Balb/c mice and lowest in DBA/2 and 129 backgrounds. Expression levels were inversely correlated with methylation levels. Similarly, Engler et al. [83] reported high transgene methylation levels in C57BL/6 mice and low or absent levels in DBA/2 mice. They were further able to map a dominant (in C57BL/6 mice) modifying allele to mouse chromosome 4 and termed this gene *Ssm-1*, for strain-specific modifier. It has been proposed that the *Ssm-1* locus acts to modify methylation by altering chromatin structure shortly after blastocyst implantation, thereby marking target loci for de novo methylation [84]. If this hypothesis is correct, the Ssm-1 protein might have a function similar to that of the polycomb proteins which act to inactivate target loci early Drosophila development. In work with the mutant strain described by Michaud et al. ([54], see *Cis*-Acting Elements as Imprinting. Signals) we have found increased expression and decreased methylation of the IAP element that has transposed into the agouti locus in the DBA/2 strain background as compared with C57BL/6 strain background. However, initial experiments have ruled out the *Ssm-1* gene as being responsible for this observation (Walsh, Turker and Bestor, unpublished). In work with an endogenous imprinted locus, Forejt and Gregorava [85] have demonstrated that the gene repression can be disrupted when crosses are made between closely related mouse subspecies.

Several recent reports have suggested roles for modifying factors in controlling methylation in genetic diseases. In one report, Smeets et al. [86] describe two brothers with the full FMR1 mutation, yet with normal phenotypes because the expanded CGG sequences in FMR1 is unmethylated. Although quite speculative, it is tempting to suggest a deficiency in this family for a factor that methylates the expanded FMR1 region. In a second report describing work with cells from patients with ICF (immunodeficiency, centromeric instability, facial abnormalities) syndrome, which is characterized at the cellular level by chromosomal fragility and hypomethylation of satellite DNA sequences, it was suggested that a *trans*-acting factor necessary to maintain chromosomal stability was absent from the ICF cells [3].

9. Conclusion

Data from a number of laboratories have suggested a role for *cis*- and *trans*-acting factors in the formation of mammalian DNA methylation patterns, although the key regulatory steps remain to be dis-

covered. Perhaps the best current model for how these factors interact to establish and maintain a DNA methylation pattern can be built for the region upstream of the mouse *aprt* gene (Fig. 2). This region includes tandem B1 repetitive elements that attract DNA methylation (methylation center) and a promoter region containing four Sp1 binding sites that repels DNA methylation. Maintenance of a methylation free promoter region (e.g., H3 site, Fig. 2) may be mediated by the presence of Sp1 protein on its bindings sites or perhaps other proteins that interact with these sites or with the Sp1 proteins. Methylation of CpG sites near the methylation center (e.g., H1 site, Fig. 2) is initiated only in primitive embryonal cell types indicating the need for one or more modifying factors and the DNA MTase protein. These factors could interact directly with the DNA MTase or directly with the repetitive sequences. Once methylation is established by the methylation center, it can spread towards the promoter region until inhibited by the presence of the methylation free region induced by the Sp1 binding sites. CpG sites located between the methylation center and the promoter (e.g., H2 site, Fig. 2) are subject to the opposing forces of the methylation center and the Sp1 binding sites, resulting in a continuous transition between methylation and demethylation that is reflected at the population level by a partial methylation profile [36].

The above speculative model is meant to describe the formation and maintenance of a methylation pattern for only a single target region in the genome, though it is hoped that it will provide a general model for the formation of methylation patterns during development. Clearly, additional data are required to determine how methylation patterns are formed during early embryonal development, the development of specific tissue types, and gametogenesis. If it is accepted that the DNA MTase and *cis*-acting elements are necessary, but not sufficient, components for methylation pattern formation in mammalian cells, then the other necessary components are modifying factors that inhibit or promote de novo and/or maintenance methylation. The identification and isolation of genes that encode these factors is required before the regulation of normal and abnormal methylation patterns in mammalian cells can truly be understood.

Acknowledgements

We thank Drs. M. Pieretti and J. Graff for critical readings of this manuscript. This work was supported by a grant from The Council for Tobacco Research (MST) and grants from NIH (THB).

References

[1] Razin, A. and H. Cedar (1993) DNA methylation and embryogenesis, in: J.P. Jost and H.P. Saluz (Eds.), DNA Methylation: Molecular Biology and Biological Significance, Birkhauser Verlag Basel/Switzerland, 344–357.
[2] Bates, G. and H. Lehrach (1994) Trinucleotide repeat expansions and human genetic disease, Bioessays, 16, 277–284.
[3] Schuffenhauer, S., O. Bartsch, M. Stumm, T. Buchholz, T. Petropoulou, S. Kraft, B. Belohradsky, G.K. Hinkel, T. Meitinger and R.-D. Wegner (1995) DNA, FISH and complementation studies in ICF syndrome: DNA hypomethylation of repetitive and single copy loci and evidence for a *trans*-acting factor, Hum. Genet., 96, 562–571.
[4] Bestor, T.H. and B. Tycko (1996) Creation of genomic methylation patterns, Nature Genet., 12, 363–367.
[5] Baylin, S.B., M. Makos, J. Wu, R.-W Yen, A. Yende Bustros, P. Vertino, P. and B.D. Nelkin (1991) Abnormal patterns of DNA methylation in human neoplasia: potential consequences for tumor progression, Cancer Cells, 3, 383–390.
[6] Bestor, T.H. (1995) The biochemical basis of allele-specific gene expression in genomic imprinting and X inactivation. in: R. Ohlsson, K. Hall and M. Ritzen (Eds.), Genomic Imprinting: Causes and Consequences, Cambridge University Press, pp. 109–116.
[7] Bird, A.P. (1995) Gene number, noise reduction and biological complexity, 1995, Trends Genet., 11, 94–100.
[8] Efstratiadis, A. (1994) Parental imprinting of autosomal mammalian genes, Curr. Opin. Genet. Dev., 4, 265–280.
[9] Migeon, B.R. (1994) X-chromosome inactivation: molecular mechanisms and genetic consequences, Trends Genet., 10, 230–235.
[10] Schmid, C.W. and W.R. Jelinek (1982) The Alu family of dispersed repetitive sequences, Science, 216, 1065–1070.
[11] Hellmann-Blumberg, U., M.F. Hintz, J.M. Gatewood and C.W. Schmid (1993) Developmental differences in methylation of human Alu repeats, Mol. Cell. Biol., 13, 4523–4530.
[12] Kochanek, S., D. Renz and W. Doerfler (1993) DNA methylation in the Alu sequences of diploid and haploid primary human cells, EMBO J., 12, 1141–1151.
[13] Rubin, C.M., C.A. VandeVoort, R.L.Tiplitz and C.W. Schmid (1994) Alu repeated DNAs are differentially methylated in primate germ cells. Nucleic Acids Res., 22, 5121–5127.
[14] Liu, W.M. and C.W. Schmid (1993) Proposed role for DNA

methylation in Alu transcriptional repression and mutational inactivation, Nucleic Acids Res., 21, 1351–1359.
[15] Li, E., C. Beard and R. Jaenisch (1993) Role for DNA methylation in genomic imprinting, Nature, 366, 362–365.
[16] Li, E., T.H. Bestor and R. Jaenisch (1992) Targetted mutation of the DNA methyltransferase gene results in embryonic lethality, Cell, 69, 915–926.
[17] Holliday, R. and J.E. Pugh (1975) DNA modification mechanisms and gene activity during development, Science, 187, 226–232.
[18] Riggs, A.D. (1975) X inactivation, differentiation, and DNA methylation, Cytogenet. Cell Genet., 14, 9–11.
[19] Taylor, J.H. (1984) Origins of replication and gene regulation, Mol. Cell. Biochem., 61, 99–109.
[20] Cleaver, J.E. (1995) It was a very good year for DNA repair, Cell, 76, 1–4.
[21] Bestor, T.H. (1990) DNA methylation: evolution of a bacterial immune function into a regulator of gene expression and genome structure in higher eukaryotes, Phil Trans. R. Soc. Lond., B326, 179–187.
[22] Doerfler, W. (1991) Patterns of DNA methylation-evolutionary vestiges of foreign DNA inactivation as a host defense mechanism, Biol. Chem. Hoppe-Seyler, 372, 557–564.
[23] Jahner, D. and R. Jaenisch (1985) Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity, Nature, 315, 594–597.
[24] Barlow, D.P. (1993) Methylation and imprinting: from host defense to gene regulation?, Science, 260, 309–310.
[25] Verdine, G.L. (1994) The flip side of DNA methylation, Cell, 76, 197–200.
[26] Bestor, T.H. (1992) Activiation of mammalian DNA methyltransferase by cleavage of Zn binding regulatory domain, EMBO J., 11, 2611–2617.
[27] Bestor, T.H. and G.L. Verdine (1994) DNA methyltransferases. Curr. Biol., 6, 380–389.
[28] Ma, Q., H. Alder, K.K. Nelson, D. Chatterjee, Y. Gu, T. Nakamura, E. Canaani, C.M. Croce, L.D. Siracusa, A.M. and Buchberg (1993) Analysis of the murine ALL-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases, Proc. Natl. Acad. Sci. USA, 90, 6350–6354.
[29] Yu, B.D., J.L. Hess, S.E. Horning, G.A.J. Brown and S.J. Korsmeyer (1995) Altered Hox expression and segmental identity in MLL-mutant mice, Nature, 378, 505–508.
[30] Leonhardt, H., A.W. Page, H.-Ul Weirer and T.H. Bestor (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei, Cell, 71, 865–873.
[31] Carlson, L.L., A.W. Page and T.H. Bestor (1992) Properties and localization of DNA methyltransferase in preimplantation mouse embryos: implications for genomic imprinting, Genes Dev., 6, 2536–2541.
[32] Trasler, J.M., D.G. Trasler, T.H. Bestor, E. Li and F. Ghibu (1996) DNA methyltransferase in normal and Dnmt1/Dnmt1n mouse embryos, Dev. Dynamics, in press.
[33] Trasler, J.M., A.A. Alcivar, L.E. Hake, T. Bestor and N.B Hecht, (1992) DNA methyltransferase is developmentally expressed in replicating and non-replicating male germ cells. Nucleic Acids Res., 20, 2541–2545.
[34] Jue, K., T.H. Bestor and J.M. Trasler (1995) Regulated synthesis and localization of DNA methyltransferase during spermatogenesis, Biol. Reprod., 53, 561–569.
[35] Silva, A.J. and R.W. White (1988) Inheritance of allelic blueprints for methylation patterns, Cell, 54, 145–152.
[36] Turker, M.S., K. Swisshelm, A.C. Smith and G.M. Martin (1989) A partial methylation profile for a CpG site is stably maintained in mammalian tissues and cell lines. J. Biol. Chem., 264, 11632–11636.
[37] Turker, M.S., P. Mummaneni and G.E. Cooper (1994) The mouse *aprt* gene as a model for epigenetic gene inactivation, Adv. Exp. Biol. Med., 370, 647–652.
[38] Kochanek, S., M. Toth, A. Dehmel, D. Renz and W. Doerfler (1990) Interindividual concordence of methylation profiles in human genes for tumor necrosis factors α and β, Proc. Natl. Acad. Sci. USA, 87, 8830–8834.
[39] Silva, A.J., K. Ward and R. White (1993) Mosaic methylation in clonal tissue, Dev. Biol., 156, 391–398
[40] Selker, E.U. (1990) DNA methylation and chromatin structure: a view from below, Trends Biochem., 15, 103–107.
[41] Singer, M.J., Marcotte, B.A. and Selker, E.U. (1995) DNA methylation associated with repeat-induced point mutation in *Neurospora crassa*. Mol. Cell. Biol., 15, 5586–5597.
[42] Szyf, M., B.P. Schimmer and J.G. Seidman (1989) Nucleotide-sequence-specific de novo methylation in a somatic murine cell line, Proc. Natl. Acad. Sci. USA, 86, 6853–6857.
[43] Hasse, A. and W.A. Schultz (1994) Enhancement of reporter gene de novo methylation by DNA fragments from the α-fetoprotein control region, J. Biol. Chem., 269, 1821–1826.
[44] Turker, M.S., P. Mummaneni and P.L. Bishop (1991) Region- and cell type-specific de novo DNA methylation in cultured mammalian cells, Somat. Cell Mol. Genet., 17, 151–157.
[45] Mummaneni, P., P.L. Bishop and M.S. Turker (1993) A *cis*-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene, J. Biol. Chem., 268, 552–558.
[46] Magewu, A.N. and P.A. Jones (1994) Ubiquitous and tenacious methylation of the CpG site in codon 248 of the *p53* gene may explain its frequent appearance as a mutational hot spot in human cancer, Mol. Cell. Biol., 14, 4225–4232.
[47] Quentin, Y. (1994) A master sequence related to a free Alu monomer (FLAM) at the origin of the B1 family in rodent genomes, Nucleic Acids Res., 25, 2222–2227.
[48] Sinnett, D., C. Richer, J.M. Deragon and D. Labuda (1991) Alu RNA secondary structure consists of two independent 7 SL RNA-like folding units, J. Biol. Chem., 266, 8675–8678.
[49] Toth, M., U., Lichtenberg, and W. Doerfler (1989) Genomic sequencing reveals a 5-methylcytosine-free domain in active promoters and the spreading of preimposed methylation patterns, Proc. Natl. Acad. Sci. USA, 86, 3728–3732.
[50] Stoger, R., P. Kubicka, C. Liu, T. Kafri, A. Razin, H. Cedar

and D.P. Barlow (1993) Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinted signal, Cell, 73, 61–71.
[51] Neumann, B., P. Kubicka and D.P. Barlow (1995) Characteristics of imprinted genes, Nature Genet., 9, 12–13
[52] Chaillet, J.R., D.S., Bader and P. Leder (1995) Regulation of gametic imprinting by gametic and embryonic processes, Genes Dev., 9, 1177–1187.
[53] Duhl, D.M.J., H. Vrieling, K.A. Miller, G.L. Wolff and G.S. Barsh (1994) Neomorphic agouti mouse mutations in obese yellow mice, Nature Genet., 8, 59–65.
[54] Michaud, E.M., M.J. van Vugt, S.J. Bultman, H.O. Sweet, M.T. Davisson and R.P. Woychik (1994) Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and influenced by parental lineage, Genes Dev., 8, 1463–1472.
[55] May-Hoopes, L.L., A. Brown and R.C.C. Huang (1983) Methylation and rearrangement of mouse in intracisternal A particle genes in development, aging and myeloma, Mol. Cell. Biol., 3, 1371–1380.
[56] Buiting, K., S. Saitoh, S. Gross, B. Dittrich, S. Schwartz, R.D. Nicholls and B. Horsthemke (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15, Nature Genet., 9, 395–400.
[57] Bestor, T. (1987) Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase, Nucleic Acids Res., 15, 3835–3843.
[58] Christman, J.K., G. Sheikhejad, C.J. Marasco and J.R. Sufrin (1995) 5-Methyl-2'-deoxycytidine in a single stranded DNA can act in *cis* to signal de novo DNA methylation, Proc. Natl. Acad. Sci. USA, 92, 7347–7351.
[59] Smith, S.S., J.L.C. Kan, D.J. Baker, B.E. Kaplan and P. Dembek (1991) Recognition of unusual DNA structures by human DNA (cytosine-5) methyltransferase, J. Mol. Biol., 217, 39–51.
[60] Oberle, I., F. Rousseau, D. Heitz, C. Kretz, D. Devys, A. Hanauer, J. Bove, M.F. Bentheas and J.L. Mandel (1991) Instability of a 550-bp DNA segment and abnormal methylation in fragileX-syndrome, Science, 252, 1097–1102.
[61] Pieretti, M., F.P. Zhang, Y.H. Fu and S.T. Warren S.T, (1991) Absence of expression of the *Fmr-1* gene in fragile X syndrome, Cell, 66, 817–822.
[62] Chen, X., S.V.S. Mariappan, P. Catasti, R. Ratliff, R.K., Moyzis, A. Laayoun, S.S. Smith, E.M. Bradbury and G. Gupta (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications, Proc. Natl. Acad. Sci. USA, 92, 5199–5203.
[63] Gacy, A.M., G. Goellner, N. Juranic, S. Macura and C.T. McMurray (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro, Cell, 81, 533–540.
[64] Nadel, Y., P. Weisman-Shomer and M. Fry (1995) The fragile X syndrome single strand d(CCG)n nucleotide repeats readily fold back to form unimolecular hairpin structures, J. Biol. Chem., 270, 28970–28977.
[65] Labuda, D., D. Sinnett, C. Richer, J.M Deragon and G. Striker (1991) Evolution of mouse B1 repeats: 7SL RNA folding pattern conserved, J. Mol. Evol., 32, 405–414.
[66] Wassenegger, M., S. Heimes, L. Ridel and H. Sanger, H (1994) RNA-directed de novo methylation of genomic sequences in plants, Cell, 76, 567–576.
[67] Penny, G.D., G.F. Kay, S.A. Sheardown, S. Rastan and N. Brockdorff (1996) Requirement for Xist in X chromosome inactivation, Nature, 379, 131–138.
[68] Razin, A. and H. Cedar (1991) DNA methylation and gene expression, Microbiol. Rev., 55, 451–456.
[69] Szyf, M., G. Tanigawa and P.L. McCarthy Jr. (1990) A DNA signal from Tdy-1 defines de novo methylation patterns in embryonic stem cells, Mol. Cell. Biol., 10, 4396–4400.
[70] Hasse, A., W.A. Schulz and H. Sies, H. (1992) De novo methylation of transfected CAT gene plasmid constructs in P9 mouse embryonal carcinoma cells, Biochim. Biophys. Acta, 1131, 16–22.
[71] Bestor, T. H., G. Gunderson, A.-B. Kolsto, and H. Prydz (1992) CpG islands in mammalian gene promoters are inherently resistant to de novo methylation, Genetic Analysis: Techniques Appl., 9, 48–53.
[72] Cooper, G.E., P. L. Bishop and M.S. Turker (1993) Hemidemethylation is sufficient for chromatin relaxation and transcriptional activation of a methylated *aprt* gene in mouse P19 embryonal carcinoma cell line, Somat. Cell. Mol. Genet., 19, 221–227.
[73] Dush, M.K., M.R. Briggs, M.E. Royce, D.A. Schaff, J.A. Tischfield and P.J. Stambrook (1988) Identification of DNA sequences required for mouse *APRT* gene expression, Nucleic Acids. Res., 16, 8509–8524.
[74] Mummaneni, P., K.A. Walker, P.L. Bishop and M.S. Turker (1995) Induction of epigenetic gene inactivation by a mammalian methylation center, J. Biol. Chem., 270, 788–792.
[75] Macleod, D., J. Charlton, J. Mullins and A. Bird (1994) Sp1 binding sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island, Genes Dev., 8, 2282–2292.
[76] Brandeis, M., D. Frank, I. Keshet, Z. Siegfried, M. Mendelsohn, A. Nemes, V. Temper, A. Razin and H. Cedar (1994) Sp1 elements protect a CpG island from de novo methylation, Nature, 371, 435–438.
[77] Jost, J.P. and Y.C. Jost (1995) Mechanism of active DNA demethylation during embryonic development and cellular differentiation in vertebrates, Gene, 157, 265–266.
[78] Jost, J. (1993) Nuclear extracts of chicken embryos promote an active demethylation of DNA excision repair of 5-methyldeoxycytidine, Proc. Natl. Acad. Sci. USA, 90, 4684–4688.
[79] Jost, J.P., M. Siegmann, L. Sun and R. Leung (1995) Mechanisms of DNA demethylation in chicken embryos, Purification and properties of a 5-methylcytosine-DNA glycosylase, J. Biol. Chem., 270, 9734–9739.
[80] Frank, D., I. Keshet, M. Shani, A. Levine, A. Razin and H. Cedar (1991) Demethylation of CpG islands in embryonic cells, Nature, 351, 239–241.
[80a] Weiss, A. et al. (1996) Cell, 86, 709–718.

[81] Chesnokov, I.N. and C.W. Schmid (1995) Specific Alu binding protein from human sperm chromatin prevents DNA methylation, J. Biol. Chem., 270, 18539–18542.
[82] Allen, N.D., M.L. Norris and M.A. Surani (1990) Epigenetic control of transgene expression and imprinting by genotype-specific modifiers, Cell, 61, 853–861.
[83] Engler, P., D. Haasch, C.A. Pinkert, L. Doglio, M. Glymer, R. Brinster and U. Storb (1991) A strain-specific modifier on chromosome 4 controls the methylation of independent transgene loci, Cell, 65, 939–947.
[84] Weng, A., T. Magnuson and U. Storb (1995) Strain-specific transgene methylation occurs early in mouse development and can be recapitulated in embryonic stem cells, Development, 121, 2853–2859.
[85] Forejt, J. and S. Gregorova (1992) Genetic analysis of genomic imprinting: an imprinter-1 gene controls inactivation of the paternal copy of the mouse Tme locus, Cell, 70, 443–450.
[86] Smeets, H.J.M, A.P.T. Smits, C.E. Verheij, J.P.G Theelen, R. Willemsen, L van de Burgt, A.T. Hoogeveen, J.C. Oosterwijk and B.A. Oostra (1995) Normal phenotype in two brothers with a full FMR1 mutation, Hum. Mol. Genet., 4, 2103–2108.

TAB 19

This material may be protected by Copyright law (Title 17 U.S. Code)

HUMAN GENE THERAPY 9:2321–2330 (November 1, 1998)
Mary Ann Liebert, Inc.

High-Resolution Analysis of Cytosine Methylation in the 5′ Long Terminal Repeat of Retroviral Vectors

LIJUN WANG, PAUL B. ROBBINS, DENISE A. CARBONARO, and DONALD B. KOHN

ABSTRACT

Retroviral vectors based on the Moloney murine leukemia virus (Mo-MuLV) are among the most commonly used vectors for stable gene transfer into mammalian cells. However, expression from the transcription unit of the Mo-MuLV long terminal repeat (LTR) has often been unsatisfactory. Transcriptional suppression of retroviral vectors *in vitro* in embryonal carcinoma (EC) cells and *in vivo* in hematopoietic stem cells (HSCs) has been associated with increased levels of cytosine methylation in the vector 5′ LTR. To obtain a comprehensive picture of the methylation pattern in the 5′ LTR of retroviral vectors, we employed the bisulfite genomic sequencing technique, which allows detection of the methylation pattern of every CpG dinucleotide in a target sequence. We studied the 5′ LTR within the Mo-MuLV-based vector, LN, and a series of multiply modified vectors, which show improved expression *in vitro* and *in vivo*. Methylation patterns of the vectors were compared in PA317 (3T3-derived) fibroblasts, which are permissive for expression from all of the vectors, and in F9 embryonal carcinoma (EC) cells, which are restrictive for expression from the parental Mo-MuLV LTR but show improved expression from the modified vectors. These analyses revealed that the levels of methylation of CpG dinucleotides were globally consistent throughout the entire LTR, including the region of transcriptional factor binding. All vectors showed no measurable methylation of CpG dinucleotides throughout the 5′ LTR in the PA317 fibroblasts. The CpG dinucleotides of the standard Mo-MuLV-based vector (LN) were highly methylated in F9 EC cells (49.1%). The doubly modified vector, MD-neo, which did not show improved expression, exhibited a relatively high level of methylation (45%), similar to that found in the LN vector. In contrast, the CpG dinucleotides of the triply modified vectors, which showed improved expression in EC cells (MND-neo and MTD-neo), were much less methylated (26.2 and 23.4%, respectively). The results extend our previous findings of an inverse correlation between gene expression and methylation of cytosine residues of the LTR of retroviral vectors.

OVERVIEW

Retroviral vectors may produce unsatisfactory expression of transgenes under transcriptional control of the Moloney murine leukemia virus (Mo-MuLV) long terminal repeats (LTRs) in embryonic stem cells and primary hematopoietic and lymphoid cells. Retroviral vectors that are not actively transcribed undergo methylation of cytosine residues in the LTR, a phenomenon often seen in the promoters of inactive genes. Vectors with multiple modifications to transcriptional control elements of the LTR show better expression than those derived directly from Mo-MuLV, in association with less methylation of cytosine residues in the LTR. Previous studies of the methylation of the retroviral LTR have used the method of Southern blot analysis of proviral DNA digested with methylation-sensitive restriction enzymes, which has limited analysis to the cytosine residues at only a few sites in the LTR. Wang *et al.* have used a method that combines selective conversion of nonmethylated cytosine residues to uracil by exposure to sodium bisulfite followed by DNA sequence analysis to determine the methylation status of every cytosine residue in the LTR of Mo-MuLV and modified vectors in murine F9 embryonal carcinoma cells. The extent of methylation of cytosine residues was consistent throughout the LTR of each vector, with increased levels of methylated cytosines found in the LTR of poorly expressed vectors. These findings add to the understanding of the relationship between gene expression and methylation

Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, and Departments of Pediatrics and Molecular Microbiology and Immunology, University of Southern California School of Medicine, Los Angeles, CA 90027.

and may help in further development of vectors that allow more effective gene expression.

INTRODUCTION

GENE THERAPY is a promising technique for treatment of a wide variety of human diseases, including genetic disorders, cancer, and AIDS. Moloney murine leukemia virus (Mo-MuLV)-based retroviral vectors are among the most commonly used vectors for gene transfer purposes because of their high efficiencies of transduction and stable chromosomal integration (Miller, 1992; Mulligan, 1993). However, expression from the transcription unit of the Mo-MuLV long terminal repeat (LTR) in transduced cells has often been unsatisfactory. A number of investigators have observed a rapid decrease in vector expression *in vivo* in various cell types, including hematopoietic stem cells (HSCs), fibroblasts, hepatocytes, and myoblasts (Kaleko *et al.*, 1990; Palmer *et al.*, 1991; Scharfmann *et al.*, 1991; Dai *et al.*, 1992; Challita and Kohn, 1994; Lund *et al.*, 1996). Our studies using serial bone marrow transfer in mice have found expression from the Mo-MuLV LTR to be severely diminished or absent in cells derived from long-lived pluripotent HSCs (Challita and Kohn, 1994). Expression from the transcription unit of the Mo-MuLV LTR has previously been shown to be inhibited in embryonal carcinoma (EC) cell lines and in embryonic stem (ES) cells, and this inactivity is accompanied by *de novo* methylation of the proviral sequences (Habers *et al.*, 1981; Jahner *et al.*, 1982). Therefore, investigations to understand the mechanisms of vector silencing and to identify vectors that will overcome silencing are of great importance to the successful application of gene therapy.

Several *cis* elements have been implicated in the transcriptional silencing of the Mo-MuLV LTR. The enhancer/promoter of Mo-MuLV has been shown to be inactive in EC and ES cells (Gorman *et al.*, 1985; Grez *et al.*, 1991). The inactivity of the enhancer is mediated by its interaction with negatively acting transcription factors (Hilberg *et al.*, 1987; Weiher *et al.*, 1987). A negative *cis* element, called the negative control region (NCR), located at the 5′ end of the LTR, has been shown to bind the transcription factor YY1, mediating transcriptional repression (Flanagan *et al.*, 1989, 1992; Becker *et al.*, 1994). Another repressive element that binds a negatively acting transcription factor (the repressor-binding protein) is located immediately downstream of the 5′ LTR, coincident with the primer-binding site (PBS) (Barklis *et al.*, 1986; Kempler *et al.*, 1993).

In previous studies by our laboratory, a series of retroviral vectors containing modifications of the Mo-MuLV transcriptional unit were constructed (Challita *et al.*, 1995). The modifications included (1) substitution of the myeloproliferative sarcoma virus (MPSV) enhancer for the Mo-MuLV enhancer, since MPSV has greater transcriptional activity than Mo-MuLV in EC cells (Hilberg *et al.*, 1987) and hematopoietic cells (Bowtell *et al.*, 1988), (2) deletion of the NCR, (3) substitution of the Mo-MuLV PBS with the PBS from the dl587rev strain, which lacks the binding site for the repressor-binding protein (Colicelli and Goff, 1987), and (4) insertion of a fragment from the murine Thy-1 promoter region, which has been reported to inhibit *de novo* methylation of adjacent heterologous sequences (Szyf *et al.*, 1990). Each of these alterations allowed greater expression in EC cells and showed additive effects when combined (Challita *et al.*, 1995; Robbins *et al.*, 1997). Specifically, two vectors containing three modifications each from the standard Mo-MuLV vector (MND-neo and MTD-neo) showed significantly improved expression in murine EC and ES cells and HSCs, whereas a vector containing only two of the modifications (MD-neo) expressed only modestly better than did the standard LN vector.

Although cytosine methylation has long been recognized as an important factor in gene silencing, its precise function and significance in gene control have remained unclear (Cedar, 1988). Previous studies in our laboratory revealed an inverse correlation between CpG methylation in the 5′ LTR and vector expression (Challita and Kohn, 1994; Challita *et al.*, 1995; Robbins *et al.*, 1997). This conclusion was drawn from investigations of methylation patterns by Southern blot analysis of a single CpG dinucleotide at the *Sma*I site in the R region of the 5′ LTR. The limited resolution of Southern blot techniques restricted the conclusions that could be drawn, since the methylation status of the CpG sites in the vicinity of *cis*-regulatory elements that bind the transcription factors could not be evaluated. Evaluation of the methylation status of such sites would be of critical interest, since it is believed that one of the mechanisms by which DNA methylation inhibits gene expression is by altering the binding of transcription factors.

To obtain a comprehensive picture of the methylation pattern in the 5′ LTR of Mo-MuLV within the LN vector and the modified vectors, we extended our previous analysis by using the bisulfite genomic sequencing technique (Clark *et al.*, 1994), which allows us to detect the methylation pattern of every CpG in a target sequence. The unmodified and modified retroviral vectors were analyzed by bisulfite genomic sequencing in F9 EC and PA317 fibroblast cells. The results reveal that the level of methylation of cytosine residues is consistent throughout the 5′ LTR and extend our previous finding that there is an inverse correlation between proviral gene expression and methylation of the 5′ LTR.

MATERIALS AND METHODS

Cell lines

PA317 amphotropic packaging cells (Miller and Buttimore, 1986) and F9 murine embryonal carcinoma cells were obtained from American Type Culture Collection (Rockville, MD). The PA317 cells, derived from NIH 3T3 fibroblasts, were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS), penicillin-streptomycin (50 units/ml), and L-glutamine (2 m*M*). The F9 EC cells were maintained on gelatin-coated tissue culture plates in DMEM supplemented with 15% FCS, penicillin-streptomycin (50 units/ml), L-glutamine (2 m*M*), with 0.1 m*M* (final) 2-mercaptoethanol.

Retroviral vectors: construction and packaging

The LN retroviral vector (Miller and Rosman, 1989) was constructed and generously provided by A. Dusty Miller (Fred Hutchinson Cancer Center, Seattle, WA). Construction of the

retroviral vectors MND-neo, MTD-neo, and MD-neo was previously described (Challita *et al.*, 1995). The derivation of the retroviral vectors and schematic representation of their 5′ LTRs are presented in Fig. 1.

A clone of PA317 packaging cells producing the LN vector at a high titer (1×10^7/ml) was provided by A. Dusty Miller. The MND-neo, MTD-neo, and MD-neo vectors were packaged by transfecting their vector plasmids into GP+E 86 cells (Markowitz *et al.*, 1988). Ecotropic viral supernatant was used to transduce exponentially growing PA317 cells in the presence of Polybrene (4 μg/ml; Sigma Chemicals, St. Louis, MO), followed by growth in G418 (0.5 mg/ml; GIBCO-BRL, Gaithersburg, MD) to generate G418-resistant PA317 pools from which amphotropic virus supernatant was collected. The amphotropic supernatants were used to transduce the target F9 cells under the same conditions as used for the PA317 fibroblasts. F9 cell pools were selected in G418 (500 μg/ml) for approximately 14 days, after which they were cultured in the absence of G418 for 1 week before harvesting for DNA extraction.

Sequencing of the proviral 5′ LTR in F9 cells

The 5′ LTR and leader region of each of the vector proviruses were sequenced to ensure correct duplication and maintenance of all modifications after packaging and serial transduction. The proviral sequences were first amplified by polymerase chain reaction (PCR) of genomic DNA extracted from transduced F9 or PA317 cells. The primers hybridized to the 5′ end of the LTR (5′-AATGAAAGACCCCACCTGTAG-3′, sense) and the 5′ end of the *neo* gene (5′-TCATAGCCGAATAGCCTCTC-

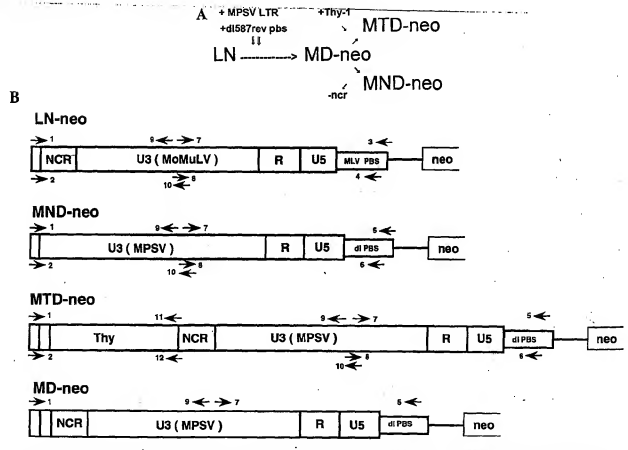


FIG. 1. Retroviral vectors (A). Derivation of the retroviral vectors. The parental vector, LN, contains the normal elements from the Moloney murine leukemia virus. The LN vector was modified by substitution of the LTR from the myeloproliferative sarcoma virus (MPSV) and replacement of the primer-binding site (PBS) region with that of the dl587rev endogenous retrovirus to produce the MD-neo vector. The MPSV LTR of MD-neo was modified by addition of a fragment from the murine Thy-1 gene to produce MTD-neo or by removal of the negative control region (NCR) to produce MND-neo. (B) Maps of long terminal repeats (LTRs) of the retroviral vectors. The LN vector contains the negative control region (NCR), U3 region from Mo-MuLV, and the Mo-MuLV primer-binding site (MLV PBS). MND-neo lacks the NCR, has the U3 region from MPSV, and has the primer-binding site from dl587rev virus (dl PBS). MTD-neo has a fragment from the murine Thy-1 gene (Thy), the U3 region from MPSV, and the primer-binding site from the dl587rev virus. MD-neo retains the NCR, has the U3 region from MPSV, and the primer-binding site from the dl587rev virus. Location of primers used for PCR amplification after bisulfite treatment and for sequencing are shown as arrows. Primers, described in Materials and Methods, are as follows: 1, 5′ LTR top 1; 2, 5′ LTR bot 1; 3, LN top 2; 4, LN bot 2; 5, MND top 2; 6, MND bot 2; 7, Down a1; 8, Down b1; 9, Down a2 re; 10, Down b1 re; 11, Thy 1 top re; 12, Thy 1 bot re.

CAC-3′, antisense). The 1.5 to 1.7-kb PCR products were gel purified and sequenced using an Applied Biosystems (Foster City, CA) 373A automated DNA sequencer.

For sequencing of the LTR enhancer regions, a primer complementary to the 5′ end of the LTR (5′-AATAAAGACCC-CACCTGTAG-3′, sense) was used. Sequencing of the PBS region was accomplished using primers from the splice donor site (5′-TGCGGTGGTCCCTGGGCAGG-3′, antisense) and sequencing of the U3 and R regions was done using a primer from the R/U5 region (5′-GAGACCACGATTCGGATGCA-3′, antisense). A primer from the Thy-1 sequence (5′-TCGGGGTGGAGCAGTC-TTCT-3′, sense) was used to sequence the MTD-neo vector containing the murine Thy-1 gene fragment.

Northern blot analysis

Total cellular RNA was isolated from F9 cells transduced by the different retroviral vectors (RNA Stat-60; Tel-Test, Friendswood, TX). RNA (15 μg) was electrophoresed on a 1.2% formaldehyde gel, denatured, neutralized, and transferred to a nylon membrane by capillary blotting. The filter was hybridized with the *neo* gene as probe, and Kodak (Rochester, NY) X-Omat AR film was exposed to the filter at −70°C for 5 days. After satisfactory exposure was obtained, the membrane was stripped, rehybridized with the mouse β-actin cDNA probe, and again analyzed by autoradiography.

Bisulfite genomic sequencing

To analyze the methylation pattern by genomic sequencing, we used a modified version of the sodium bisulfite method of Frommer *et al.* (1992). Genomic DNA was cut into fragments with *Bam*HI and *Bgl*II (GIBCO-BRL), which do not cut within any of the vectors, ethanol precipitated, and quantified using a TKO 100 fluorometer (Hoefer Instruments, San Francisco, CA). Restriction-digested DNA samples (2 μg in 20 μl of water) were denatured by adding freshly prepared 3 *N* NaOH to a final concentration of 0.6 *M* and incubating for 20 min at 45°C. Alkali-denatured DNA samples were incubated in 3.1 *M* sodium bisulfite (Sigma) pH 5.0, at a DNA concentration of 8.3 μg/ml for 16 hr at 40–50°C. The bisulfite-treated DNA samples were recovered using the Wizard DNA clean-up system (Promega, Madison, WI). The eluted DNA was alkali denatured in 0.27 *N* NaOH for 10 min at 37°C, ethanol precipitated, and resuspended in DNA buffer (10 m*M* Tris-HCl, [pH 8.0], 0.1 m*M* EDTA).

Bisulfite-treated DNA samples were PCR amplified using pairs of strand-specific primers. Fifty-microliter PCR reactions were performed in buffer containing 10 m*M* Tris-HCl, (pH 8.3), 50 m*M* KCl, 3 m*M* $MgCl_2$, 0.2 m*M* deoxynucleoside triphosphates, a 1 μ*M* concentration of each primer, and 2 U of *Taq* DNA polymerase (Perkin-Elmer, Norwalk, CT). The samples were subjected to 2 min of denaturation at 94°C, 30 cycles of 94°C (1 min), 65°C (2 min), 72°C (2 min), 72°C (1.5 min), and a final step of 6 min of extension at 72°C.

The primers used for strand-specific amplification of bisulfite-treated DNA were

5′ LTR top 1: 5′-TGAAAGATTTATTTGTAGGTTTG-GTAAGT-3′
5′ LTR bot 1: 5′-AATAAGAAACCCCACCTATACATT-TAAC-3′
LN top 2: 5′-CATAACATCAAACATAAACACTAAACA-AT-3′
LN bot 2: 5′-ATAATTGGATAGATATAGATAAGTTGT-TGG-3′
MND top 2: 5′-AATCAATCTAAAAAAACCCTCCCAA-AA-3′
MND bot 2: 5′-AGGGGTTTTTAAATTTTGGTGGGA-3′

The primers 5′ LTR top 1 and LN top 2 were used to produce a 639-bp fragment from the upper strand of the LN vector. The primer pair 5′ LTR bot 1 and LN bot 2 were used to produce a 617-bp fragment from the bottom strand of the LN vector. A 574-bp fragment from the top strand of the MND vector was generated using primers 5′ LTR top 1 and MND top 2. A 563-bp fragment from the bottom strand of MND vector was amplified using primers 5′ LTR bot 1 and MND bot 2. Primers 5′ LTR top 1 and MND top 2 were used to generate an 845-bp fragment from the upper strand of the MTD vector. Primers 5′ LTR bot 1 and MND bot 2 were used to amplify an 834-bp fragment from the bottom strand of the MTD vector. A 636-bp fragment from the top strand of the MD vector was amplified with primers 5′ LTR top 1 and MND top 2.

PCR products were gel purified, and sequenced directly to analyze the average level of methylation in populations of transduced cells. The sequencing reactions were performed according to the ^{33}P terminator cycle sequencing kit (Amersham, Arlington Heights, IL) and analyzed on denaturing polyacrylamide gels. Kodak X-Omat AR film was exposed to vacuum-dried gels over 1–3 days at room temperature.

The internal primers used for sequencing were

Down a 1: 5′-TTAGGGTGTTTTAAGGATTTGAAATGA-TTT-3′
Down b1: 5′-ATTTCCAAAATACCCCAAAAACCTAAA-ATA-3′
Down a2 re: 5′-ATCTAATAATTCTCTAAAAACTACTAA-AAA-3′
Down b1 re: 5′-TATTTTAGGTTTTTGGGGTATTTTG-GAAAT-3′
Thy top re: 5′-AAACTAAAAAAAATATAAATAAAAAAC-TAC-3′
Thy bot re: 5′-GGGTATGGGTAGAAGATTGTTTTATTT-3′

RESULTS

Expression by the series of vectors in the F9 embryonal carcinoma cells was assessed by Northern blot analysis. The level of transcripts from the LN vector in F9 cells was below the limit of detection (Fig. 2), even though the cells had been selected in G418. Expression from the doubly modified vector, MD-neo, was low but detectable. Expression by both of the triply modified vectors, MND-neo and MTD-neo, was markedly higher, with transcripts readily seen by Northern analysis. It should be noted that the level of expression by the LN vector in 3T3 fibroblast-based PA317 cells greatly exceeded that of any of the vectors in F9 cells. We have previously observed a similar pattern of expression by these vectors in these cell types (Challita *et al.*, 1995).

To explore the relationship between proviral methylation and gene expression, we analyzed the methylation state of the CpG

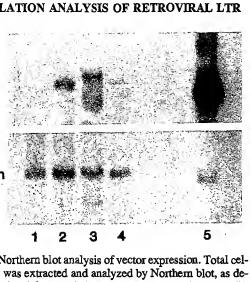


FIG. 2. Northern blot analysis of vector expression. Total cellular RNA was extracted and analyzed by Northern blot, as described in Materials and Methods. Samples were 15 μg of RNA from F9 cells transduced by the LN vector (lane 1), the MND-neo vector (lane 2), the MTD-neo vector (lane 3), and the MD-neo vector (lane 4). RNA (5 μg) from PA317 cells transduced by the LN vector is shown in lane 5. *Top*: Results with the *neo* gene probe. *Bottom*: Results with the murine β-actin cDNA probe.

dinucleotides within the 5′ LTR of the LN, MND-neo, MTD-neo, and MD-neo retroviral vectors, using the bisulfite genomic sequencing technique (Clark *et al.*, 1994). Vectors were studied in PA317 murine 3T3 fibroblasts, which are permissive for expression, and in F9 murine EC cells, which are restrictive for expression.

The most critical element in the use of bisulfite sequencing is the design of PCR primers for the upper and lower DNA strands that can function after bisulfite conversion. Even with the aid of computerized PCR primer design algorithms, much of the primer identification is empirical. Primers must be designed to be complementary to the sequences expected to be present after bisulfite conversion of all cytosines to uracils. Because bisulfite treatment results in DNA sequences that are A–T rich, identification of unique primer sequences without intra- and interprimer complementarity is difficult. Because bisulfite treatment results in the upper and lower strands of DNA becoming noncomplementary at nonmethyl-C/G base pairs, different primers are needed to amplify each strand, even though they span the same region. Primers should be chosen from sequences that do not contain CpG dinucleotides, because the potential for variable levels of cytosine methylation at these sites makes it difficult to predict the identity of the bases remaining after bisulfite conversion.

Application of the bisulfite sequencing method to the specific task of determining the methylation status of cytosine residues in the 5′ LTR of Mo-MuLV retroviral vectors poses some further, unique obstacles. Because the 5′ LTR of the vector is the site of transcriptional initiation, its methylation pattern is likely to be of more relevance than that of the 3′ LTR. The presence of identical sequences in both the 5′ and 3′ LTRs requires that one PCR primer be located at vector sequences downstream from the 5′ LTR, to amplify it speeifically. The need to span the entire length of the LTR and beyond necessitates amplification of relatively long sequences (>600–800 bp), which is inefficient using DNA that has been partially degraded from the bisulfite treatment protocol. In addition, the presence in the murine cell genome of multiple endogenous retroviral sequences with homology to the LTR of the vector may increase background amplification.

In spite of these obstacles, functional primers were created for the upper and lower strands of each vector and used to conduct the assay. For all samples in each set, the bisulfite conversion reactions, PCR amplification, and subsequent sequencing reactions were carried out in parallel. Bisulfite-induced C → T transitions (G → A on the lower strand), which indicate a lack of methylation, were quantitated by densitometer.

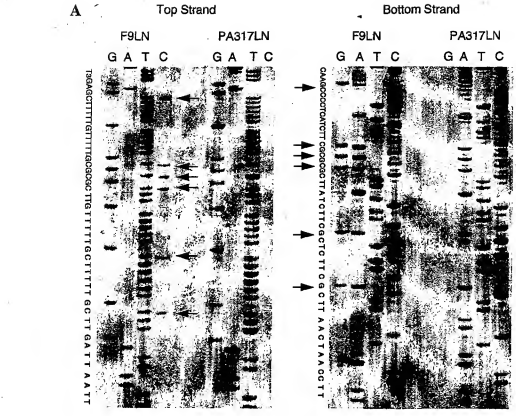
Figure 3 shows representative blots that highlight bisulfite conversion of unmethylated cytosines to uracil. The intensities of the cytosine bands of the upper strand decrease with conversion to thymine. On the bottom strand of the proviral DNA, the conversion is detected as a transition of guanine residues to adenine. This effect is clearly apparent in Fig. 3, by comparing the intensity of the bands in the cytosine lane of the sequences of the LN vector in F9 cells, compared with PA317 cells. Because PA317 cells do not methylate exogenous retroviral sequences to any significant extent, all of the cytosines were converted to uracil and, consequently, were in the thymine lane of the sequencing gel. In contrast, since the F9 cells heavily methylate Mo-MuLV vector DNA, the methylated cytosines were not converted to uracil and, thus, persisted in the cytosine lane.

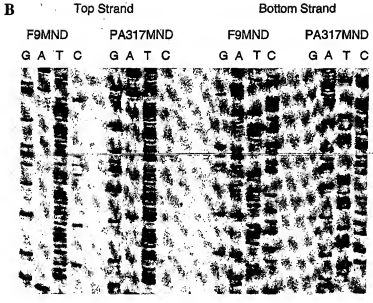
From among the 19 CpG sites analyzed in the 5′ LTR of the LN vector provirus in F9 EC cells, the extent of methylation of cytosine residues ranged between 36 and 70%, with an average value of 49.1%. The cytosines in the upper strand of the 5′ LTR were 50.6% methylated on average, while the lower strand showed an average of 47.5% methylation (Fig. 4A).

The methylation of cytosines in the MND-neo vector was determined in PA317 and F9 cells. As expected, in PA317 cells, all of the cytosines in the MND-neo 5′ LTR were converted to thymine (Fig. 3B), leaving no detectable bands in the cytosine lane of the sequencing gel. The extensive conversion from C → T in the upper strand and G → A in the lower strand indicates that the MND-neo vector was less methylated in F9 cells than was the LN vector (Fig. 4B). Densitometric analysis of 21 CpG sites showed that the extent of methylation of cytosine residues ranged between 12 and 34%, with an average value of 26.2%. The cytosines in the upper strand of the MND-neo 5′ LTR were 25.7% methylated, while the lower strand showed an average value of 26.7% methylation (Fig. 4B). In general, the extent of methylation was similar for opposing cytosine residues of CpG dinucleotides of opposing strands.

Analysis of the methylation of cytosines of the 5′ LTR of the MTD-neo vector in F9 cells showed results similar to those seen with MND-neo. The extent of cytosine methylation in the 5′ LTR of the MTD-neo vector ranged from 10 to 36%, with an average of 23.4% methylation (Fig. 4C). The cytosines of the upper strand of the MTD-neo vector were 24.2% methylated and those of the lower strand were 22.6% methylated.

In contrast, MD-neo, which retains the NCR sequences and does not have the Thy-1 fragment, was methylated to an extent similar to that of the Mo-MuLV-based LN vector (Fig. 3C). Densitometric analysis revealed that the cytosine residues in the upper strand of the MD-neo vector 5′ LTR were methylated from 25 to 61%, with an average of 45% methylation (Fig. 4D).





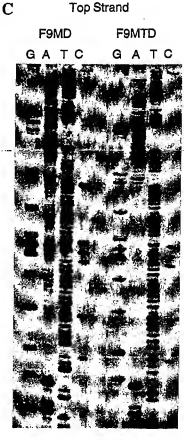


FIG. 3. Sequencing gels from bisulfite-treated DNA. (A) LN vector in F9 and PA317 cells; (B) MND-neo vector in F9 and PA317 cells; (C) MD-neo and MTD-neo vectors in F9 cells.

The status of the lower strand could not be determined, despite evaluation of multiple sets of primers.

DISCUSSION

Retroviral vectors have become a mainstay of gene therapy efforts, especially for applications where stable persistence of the transferred gene is needed in cells that undergo significant proliferation, such as T lymphocytes and hematopoietic stem cells. Persistent expression of the transferred gene is also required to allow long-term benefits to be realized. Expression of genes in retroviral vectors may be regulated from internal promoters that have "housekeeping" properties, enabling them to remain active under a variety of cellular activation and differ-

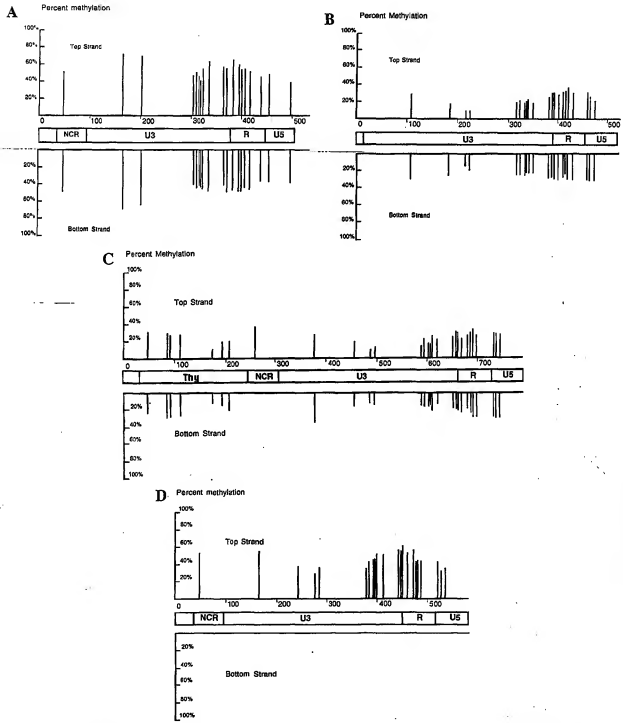


FIG. 4. Percent methylation measured at each CpG dinucleotide of retroviral vectors in F9 cells. (A) LN vector; (B) MND-neo vector; (C) MTD-NEO vector; (D) MD-neo vector.

entiation states, e.g., the phosphoglycerate kinase promoter (Apperley *et al.*, 1991). However, simple retroviral vectors in which the gene of interest is under transcriptional control of the 5′ LTR typically can be produced at higher titers than more complicated vectors. While the Mo-MuLV LTR is capable of high levels of expression of genes under its control in proliferating cells in culture, poor expression from the LTR has been observed *in vivo* in hepatocytes, keratinocytes, myocytes, pluripotent murine hematopoietic stem cells, and quiescent murine and human T lymphocytes (Palmer *et al.*, 1991; Scharfmann *et al.*, 1991; Dai *et al.*, 1992; Challita and Kohn, 1994; Agarwal *et al.*, 1998). Understanding and overcoming the mechanisms that lead to poor or inconsistent expression from the Mo-MuLV LTR is, therefore, important in developing effective gene therapy approaches to a variety of genetic, malignant, and infectious diseases.

Methylation of cytosine residues in DNA (usually in CpG dinucleotides) has been found in association with decreased gene expression in many circumstances (Cedar, 1988). Specifically, increased levels of cytosine methylation in the LTR of retroviruses (both wild-type and recombinant vectors) have been observed in a number of cell types where expression is absent, including embryonic carcinoma and embryonic stem cells (Jahner *et al.*, 1982; Challita *et al.*, 1995; Robbins *et al.*, 1997), fibroblasts (Hoeben *et al.*, 1991), colonic carcinoma cells (Lengauer *et al.*, 1997), and hematopoietic stem cells (Challita and Kohn, 1994; Robbins *et al.*, 1998). While the association between methylation and expression inactivity is a constant finding, the direction of causality remains unknown. Methylcytosines may act as a nucleus for binding of proteins (e.g., MECP-1), which then recruit the region of DNA into inactive heterochromatin (Boyes and Bird, 1991); alternatively, DNA that is first transcriptionally inactive may subsequently become subject to methylation. The specific sites of methylation may also be important, as methylation of the DNA sequences recognized by some transcriptional activating proteins may preclude their binding. Therefore, a complete analysis of the methylation status of cytosine residues throughout the 5′ LTR may provide insight into key sites where methylation is more extensive in nonexpressing DNA or that are coincident with known transcriptional factor-binding sites.

To provide a complete assessment of cytosine methylation of the 5′ LTR of retroviral vectors, we used the bisulfite sequencing method, which can determine the relative extent of cytosine methylation at each residue. Primers were developed that were able to amplify the products of bisulfite-treated DNA from the 5′ LTR of standard Mo-MuLV-based vectors and a series of modified vectors. Methylation was examined for the vectors in PA317 murine fibroblasts, where all the vectors show strong expression, and in F9 murine embryonal carcinoma cells, where expression is poor from the Mo-MuLV-based vector, LN, and a doubly modified vector, MD-neo, and expression is high from the triply modified vectors, MND-neo and MTD-neo (Challita *et al.*, 1995; Robbins *et al.*, 1998).

Using the approach described here, we were able to determine the methylation status of the cytosine residues of the 5′ LTR of each of the vectors. We found that for each vector in a specific cell type, the extent of methylation was relatively consistent for all cytosine residues in CpG dinucleotides throughout the LTR. Specifically, the methylation levels of the CpG dinucleotides in the regions of the LTR enhancer repeats were similar to those found in upstream and downstream sites. These results, therefore, corroborate the previous findings, which analyzed methylation at the single *SmaI* site by Southern blot analysis (Challita *et al.*, 1995).

The patterns of methylation were essentially symmetrical for the top and bottom strands of all of the vectors. Hemimethylated DNA is a powerful inducer of the maintenance function of the 5-methylcytosine DNA methyltransferase, which methylates cytosine residues of newly synthesized DNA strands if the opposing CpGs are methylated. Therefore, we would expect to see symmetrical methylation patterns after genomic methylation is established. These results provide direct evidence of this phenomenon in retroviral vectors.

The correlation between poor expression and increased methylation of these vectors, which we previously measured at the *SmaI* site, was seen on analysis of all cytosine residues using the bisulfite sequencing analysis. All of the vectors showed no detectable methylation of CpG sites in the PA317 murine fibroblast cell line, where all show strong expression from the LTR. In the F9 murine EC cells, we found relatively high global methylation of cytosines in the 5′ LTR of the poorly expressed vectors LN and MD-neo, and lower methylation for the vectors MND-neo and MTD-neo, which show better expression.

Our results presented here, based on bisulfite sequencing, are in accord with our previous findings, based on Southern blot analysis of the *SmaI* site, that the 5′ LTR of the triply modified vectors MND-neo and MTD-neo are less susceptible to methylation than the relatively similar, doubly modified MD-neo vector (Challita *et al.*, 1995; Robbins *et al.*, 1998). All three of these vectors contain the enhancer from the MPSV virus LTR substituting for that from Mo-MuLV, and the primer-binding site from the dl587rev virus, which does not bind the repressor protein-binding site, replacing the Mo-MuLV primer-binding site. The MPSV enhancers have Sp1-binding sites, not present in the Mo-MuLV enhancers, which have been shown to be responsible for improved expression in embryonic stem cells, compared with that from the Mo-MuLV LTR (Grez *et al.*, 1991). Binding of the Sp1 transcriptional factor is relatively insensitive to methylation and the presence of Sp1 sites has been implicated in preventing regional methylation (Harrington *et al.*, 1988; Brandeis *et al.*, 1994). Nevertheless, the presence of the MPSV enhancer (and the substituted primer-binding site) in MD-neo was not sufficient to decrease methylation or allow increased methylation.

The additional modifications made to the MD-neo vector to produce MND-neo and MTD-neo *did* result in decreased LTR methylation. The nature of the changes was quite different for the two vectors, in that MND-neo has a deletion of 66 bp near the 5′ end of the LTR (bases 33–96), whereas MTD-neo has an insertion at the same site in the LTR (base 33) of a 214-bp fragment from the murine Thy-1 gene 5′ flanking region. The deletion of the LTR in MND-neo removes a negative control region (NCR) that contains binding sites for the transcriptional factor YY1, which may suppress expression. The observations with MND-neo suggest that alleviation of transcriptional repression by removal of the NCR may prevent methylation from being imposed. The Thy-1 fragment inserted to generate MTD-neo has been previously shown to be able to prevent regional methylation (Szyf *et al.*, 1990). The decreased methylation and

improved expression activity of the MTD-neo vector suggest that methylation is the primary event, which then causes transcriptional inactivity.

Thus, the findings with these two vectors are contradictory and do not definitively discriminate the direction of causality between transcriptional inactivity and DNA methylation. In fact, the contrary findings suggest that either process may influence the other. A common mechanism may be mediated through effects on regional chromatin structure, which may be induced into an active, open state by the presence of transcriptional factors or induced into a closed, inactive state by methylation of cytosine residues. The dynamic interplay among these elements may then determine the net result of gene expression. Ongoing studies to examine the chromatin structure at the integrated LTR of these vectors may provide further insight into these processes.

ACKNOWLEDGMENTS

These studies were supported by grants from the Department of Health and Human Services (1PO1 CA59318 and 1RO1 DK49000). D.B.K. is the recipient of an Elizabeth Glaser Scientist Award from the Pediatric AIDS Foundation.

REFERENCES

AGARWAL, M., AUSTIN, T.W., MOREL, F., CHEN, J., BÖHNLEIN, E., and PLAVEC, I. (1998). Scaffold attachment region-mediated enhancement of retroviral expression in primary T cells. J. Virol. **72**, 3720–3728.

APPERLEY, J.F., LUSKEY, B.D., and WILLIAMS, D.A. (1991). Retroviral gene transfer of human adenosine deaminase in murine hematopoietic cells: Effect of selectable marker sequences on long-term expression. Blood **78**, 310–317.

BARKLIS, E., MULLIGAN, R.C., and JAENISCH, R. (1986). Chromosomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells Cell **47**, 391–399.

BECKER, K.G., JEDLICKA, P., TEMPLETON, N.S., LIOTTA, L., and OZATO, K. (1994). Characterization of hUCRBP (YY1, NF-E1, delta), a transcription factor that binds the regulatory regions of many viral and cellular genes. Gene **150**, 249–266.

BOWTELL, D.D., CORY, S., JOHNSON, G.R., and GONDA, T.J. (1988). Comparison of expression in hematopoietic cells by retroviral vectors carrying two genes. J. Virol. **62**, 2464–2473.

BOYES, J., and BIRD, A. (1991). DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell **64**, 1123–1134.

BRANDEIS, M., FRANK, D., KESHET, I., SIEGFRIED, Z., MENDELSOHN, M., NEMES, A., TEMPER, V., RAZIN, A., and CEDAR, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature (London) **371**, 435–438.

CEDAR, H. (1988). DNA methylation and gene activity: Minireview. Cell **53**, 3–4.

CHALLITA, P.M., and KOHN, D.B. (1994). Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation. Proc. Natl. Acad. Sci. U.S.A. **91**, 2567–2571.

CHALLITA, P.M., SKELTON, D., EL-KHOUEIRY, A., YU, X.M., WEINBERG, K.I., and KOHN, D.B. (1995). Multiple modifications in *cis*-elements of the LTR of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma (EC) cells. J. Virol. **69**, 748–755.

CLARK, S.J., HARRISON, J., PAUL, C.L., and FROMMER, M. (1994). High sensitivity mapping of methylated cytosines. Nucleic Acids Res. **22**, 2990–2997.

COLIGELLI, J., and GOFF, S.P. (1987). Isolation of a recombinant murine leukemia virus utilizing a new primer tRNA. J. Virol. **57**, 37–45.

DAI, Y., ROMAN, M., NAVIAUX, R.K., and VERMA, I.M. (1992). Gene therapy via primary myoblasts: Long-term expression of factor IX protein following transplantation in vivo. Proc. Natl. Acad. Sci. U.S.A. **89**, 10892–10895.

FLANAGAN, J.R., KRIEG, A.M., MAX, E.E., and KHAN, A.S. (1988). Negative control region at the 5' end of the murine leukemia virus long terminal repeats. Mol. Cell Biol. **9**, 739–746.

FLANAGAN, J.R., BECKER, K.G., ENNIST, D.L., GLEASON, S.L., DRIGGERS, P.H., LEVI, B.Z., APPELLA, E., and OZATO, K. (1992). Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. Mol. Cell Biol. **12**, 38–44.

FROMMER, M., MCDONALD, L.E., MILLAR, D.S., COLLIS, C.M., WATT, F., GRIGG, G.W., MOLLOY, P.L., and PAUL, C.L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proc. Natl. Acad. Sci. U.S.A. **89**, 1827–1831.

GORMAN, C.M., RIGBY, P.W., and LANE, D.P. (1985). Negative regulation of viral enhancers in undifferentiated embryonic stem cells. Cell **42**, 519–526.

GREZ, M., ZORNIG, M., NOWOCK, J., and ZIEGLER, M. (1991). A single point mutation activates the Moloney murine leukemia virus long terminal repeat in embryonic stem cells. J. Virol. **65**, 4691–4698.

HABERS, K., SCHNIEKE, A., STUHLMAN, H., JAHNER, D., and JAENISCH, R. (1981). DNA methylation and gene expression: Endogenous retroviral genome becomes infectious after molecular cloning. Proc. Natl. Acad. Sci. U.S.A. **78**, 7609–7613.

HARRINGTON, M.A., JONES, P.A., IMAGAWA, M., and KARIN, M. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. Proc. Natl. Acad. Sci. U.S.A. **85**, 2066–2070.

HILBERG, F., STOCKING, C., OSTERTAG, W., and GREZ, M. (1987). Functional analysis of a retroviral host-range mutant: Altered long terminal repeat sequences allow expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. U.S.A. **84**, 5232–5236.

HOEBEN, R.C., MIGCHIELSEN, A.A.J., VAN DER JAGT, R.C.M., VAN ORMONDT, H., and VAN DER EB, A.J. (1991). Inactivation of the Moloney murine leukemia virus long terminal repeat in murine fibroblasts cell lines is associated with methylation and dependent on its chromosomal position. J. Virol. **65**, 904–912.

JAHNER, D., STUHLMANN, H., STEWART, C.L., HABERS, K., LOHLER, J., SIMON, I., and JAENISCH, R. (1982). De novo methylation and expression of retroviral genomes during mouse embryogenesis. Nature (London) **298**, 623–628.

KALEKO, M., GARCIA, J.V., OSBORNE, W.R.A., and MILLER, A.D. (1990). Expression of human adenosine deaminase in mice after transplantation of genetically-modified bone marrow. Blood **75**, 1733–1741.

KEMPLER, G., FREITAG, B., BERWIN, B., NANASSY, O., and BARKLIS, E. (1993). Characterization of the Moloney murine leukemia virus stem cell-specific repressor binding site. Virology **193**, 690–699.

LENGAUER, C., KINZLER, K.W., and VOGELSTEIN, B. (1997). DNA methylation and genetic instability in colorectal cancer cells. Proc. Natl. Acad. Sci. U.S.A. **94**, 2545–2550.

LUND, A.H., DUCH, M., and PEDERSEN, F.S. (1996). Transcriptional silencing of retroviral vectors. J. Biomed. Sci. **37**, 365–378.

MARKOWITZ, D., GOFF, S., and BANK, A. (1988). A safe packag-

ing line for gene transfer: Separating viral genes on two different plasmids. J. Virol. **62,** 1120–1124.

MILLER, A.D. (1992). Retroviral vectors. Curr. Top. Microbiol. Immunol. **158,** 1–24.

MILLER, A.D., and BUTTIMORE, C. (1986). Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol. Cell Biol. **6,** 2895–2902.

MILLER, A.D., and ROSMAN, G.J. (1989). Improved retroviral vectors for gene transfer and expression. Biotechniques **7,** 980–990.

MULLIGAN, R.C. (1993). The basic science of gene therapy. Science **260,** 926–932.

PALMER, T.D., ROSMAN, G.J., OSBORNE, W.R., and MILLER, A.D. (1991). Genetically modified skin fibroblasts persist long after transplantation but gradually inactivate introduced genes. Proc. Natl. Acad. Sci. U.S.A. **88,** 1330–1334.

ROBBINS, P.B., YU, X.J., SKELTON, D.M., PEPPER, K.A., WASSERMAN, R.M., ZHU, L., and KOHN, D.B. (1997). Increased probability of expression from modified retroviral vectors in embryonal stem cells and embryonal carcinoma cells. J. Virol. **71,** 9466–9474.

ROBBINS, P.B., SKELTON, D.M., YU, X.J., HALENE, S.A., LEONARD, E.H., and KOHN, D.B. (1998). Consistent, persistent expression from modified retroviral vectors in murine hematopoietic stem cells. Proc. Natl. Acad. Sci. U.S.A. **95,** 10182–10187.

SCHARFMANN, R., AXELROD, J.H., and VERMA, I.M. (1991). Long-term in vivo expression of retrovirus-mediated gene transfer in mouse fibroblast implants. Proc. Natl. Acad. Sci. U.S.A. **88,** 4626–4630.

SZYF, M., TANIGAWA, G., and McCARTHY, P.L., Jr. (1990). A DNA signal from the Thy-1 gene defines *de novo* methylation patterns in embryonic stem cells. Mol. Cell Biol. **10,** 4396–4400.

WEIHER, H., BARKLIS, E., OSTERTAG, W., and JAENISCH, R. (1987). Two distinct sequence elements mediate retroviral gene expression in embryonal carcinoma cells. J. Virol. **61,** 2742–2746.

Address reprint requests to:
Dr. Donald B. Kohn
Mailstop #62
Childrens Hospital Los Angeles
4650 Sunset Blvd.
Los Angeles, CA 90027

Received for publication April 8, 1998; accepted after revision August 5, 1998.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Application of:	Gregory T. Bleck, et al.	
Serial No.:	10/759,315	Group No.: 1633
Filed:	1/16/04	Examiner: POPA

Entitled: **PRODUCTION OF HOST CELLS CONTAINING MULTIPLE INTEGRATING VECTORS BY SERIAL TRANSDUCTION**

FIFTH DECLARATION OF DR. GREGORY BLECK

I, Dr. Gregory Bleck, state as follows:

1. My present position is Senior Director, Cell Line Engineering, Catalent Pharma Solutions.

2. I am an inventor of the above referenced patent application.

3. At page 7 of the Office Action, the Examiner recognizes that Mathor and Burns do not teach serial transduction to obtain cells comprising genomes with 20 to 100 integrated vectors. The Examiner then goes on to present arguments as why a person of skill in the art would use serial transduction to obtain cells comprising genomes with 20 to 100 integrated retroviral vectors. The Examiner supports this argument by citations to several publications. However, with one exception, the Examiner has made assumptions that are not factually supported by those references. The one exception is that the references do teach that there is a positive correlation between MOI and integration events. Many of the other assumptions made by the Examiner overstate what the references teach, especially in relation to the invention.

4. At page 7, the Examiner states that Mathor et al. do teach that protein expression is directly proportional to integration events (i.e., copy number)(p. 10376, column 1). The Examiner goes on to state that:

> "It would have been obvious to one of skill in the art, at the time the invention was made, to modify the method of Mathor et al. and Burns et al. by serially transducing their cells with high MOIs (such as MOIs of 1,000) to achieve the claimed ranges of integration

events, with a reasonable expectation of success. The motivation to do so is provided by Mathor et al., who teach the possibility of specifying the level of transgene expression by controlling the integration events (Abstract, p. 10376, column 1).

The Examiner cites the abstract and p. 10376 of Mathor. Mathor et al. state in this section that "The rate of secretion of the exogenous protein by cultures generated by single clones was proportional to the number of integrations per progenitor cell."

5. The Examiner has overstated the commentary and data in Mathor et al.. It is a fact that Mathor et al. does not contain any data or any statement that the level of transgene expression can be controlled by controlling integration events with the range of 20 to 100 integrations. Mathor et al. presents the data on proviral integration and transgene expression on p. 10373 and in Table 1. This data shows increasing transgene expression as the proviral integrations increase from 1 to 8. When the number of proviral integrations increases to 15, the transgene expression is actually decreases to a level lower than was observed with 8 integrations. Thus, as a factual matter, Mathor et al. teaches that transgene expression correlates with number of integrations over the range of 1 to 8 integrations. Transgene expression decreased when a cell line with 15 integrations was analyzed. As a result, Mathor's statement that "The rate of secretion of the exogenous protein by cultures generated by single clones was proportional to the number of integrations per progenitor cell" is valid with respect to the range of 1 to 8 integrations and does not apply outside of that range. The Examiner's attempt to apply the statement outside of the range is not factually supported, i.e., supported by the data.

6. None of the other references relied on by the Examiner teach a correlation of transgene expression to integration number in the claimed range. Schott et al. teaches a correlation over the range of 1 to 9 integrations. See p. 304, Fig. 9. The increase between 4 and 5 integrations and 6 and 7 integration is much greater than the increase between 7 and 9 integrations. This is similar to the Mathor et al. data and indicates that transgene expression levels off as opposed to continuing to increase, although, as explained in more detail below, predications outside of the data range cannot be validly made.

7. I have addressed the data in Mathor et al. in my previous Declaration. The Examiner addresses the previous Declaration at pages 13-14 of the Office Action. The Examiner states:

> The applicant argues that the fact that different clones can produce different amounts of protein has no relevance to whether a person of skill in the art would modify Mathor and make clones with 20 or more integrated retroviral vectors. This is not found persuasive. On the contrary, such knowledge in the art does have relevance to whether one of skill in the art would make clones with 20 or more integrated retroviral vectors. The argument that the data in Table 1 of Mathor et al., which is limited to a maximum of 15 integrations cannot be extrapolated to a situation where there are 20 integrations is just an argument not supported by any evidence. Based on the teachings in the prior art (including Liu, Stamps, and Mathor et al.), one of skill in the art would have known that protein production is proportional to the number of integrated copies and that retroviral insertion is random and that expression level is dependent on the insertion sites; therefore, one of skill in the art would not conclude that the data in Table 1 indicates a maximum of 15 integrations. Based on the teachings in the art as a whole, one of skill in the art would have had reasonably expected that clones comprising more than 15 integrations would express higher amounts of protein and would have known to look for several clones having higher integration numbers and select the high producer clones.

The Examiner states that the argument that the data in Table 1 of Mathor et al., which is limited to a maximum of 15 integrations cannot be extrapolated to a situation where there are 20 integrations is just an argument not supported by any evidence. This is not true. First, the data in Table 1 of Mathor et al. show that a clone with 15 integrations has a lower level of expression than a clone with 8 integrations. This is a fact. It is also a fact that the experiments in Mathor et al. were not conducted in a manner so that a statistical analysis could be conducted. The groups were not replicated and there is no way to determine experimental error. Thus, it is not possible to construct a curve or equation from the data so that a correlation of transgene expression to a number of integrations outside of the data range (i.e., 20 to 100 integrations) can be made. Any attempt to do so is speculation without a factual basis. For example, based on the data in Mathor, it is speculation as to whether another clone with 15 integrations would have a level of transgene expression that is higher or lower than the reported clone. The reason for this is that the data is not amenable to statistical analysis so that such a prediction can be made.

This is the reason why the Examiner is incorrect in arguing that "one of skill in the art would have had reasonably expected that clones comprising more than 15 integrations would express higher amounts of protein and would have known to look for several clones having higher integration numbers and select the high producer clones." The data in Mathor et al. is not amenable to this assumption. Because the experiments were not replicated and because there is not data on multiple clones with 15 (or a similar number) of integrations, it is not possible to predict or comment on the amount of expression one could expect from another clone with 15 integrations. The single clone reported in Mathor et al. could be an example of the upper limit of expression, the median level of expression, or a low level of expression. The fact is, without additional data, one cannot know whether other clones with 15 integrations would have higher levels of expression than the observed clone. All that one knows for sure is the fact that the Mathor et al. data shows that expression from the clone with 15 integrations was lower than the expression observed in the clone with 8 integrations.

The Examiner further states that based on the teachings in the prior art (including Liu, Stamps, and Mathor et al.), one of skill in the art would have known that protein production is proportional to the number of integrated copies and that retroviral insertion is random and that expression level is dependent on the insertion sites; therefore, one of skill in the art would not conclude that the data in Table 1 indicates a maximum of 15 integrations.

As explained above, the data in Mathor et al. indicates that transgene expression increases with increases number of integrations up to 8 integrations. Transgene expression decreased when the number of integrations increased to 15. No other conclusions can be made based on this data. Furthermore, Lui et al. contains data on the correlation of expression of transgenes separated by an IRES and does not address transgene expression correlated to number of integrations. See Abstract, Fig. 2, Fig. 4, Fig. 5. Stamps et al. examined the role of the T-antigen gene and its site of integration in human epithelial cell immortalization. p. 871, Col. 2, first full para. The cells examined had up to five integrations. See Fig. 2 and Fig. 3. Stamps et al. does not comment on a correlation of transgene expression to number of integrations. These references do not provide factual support for the Examiner's argument.

8. In my previous Declaration, I provided evidence showing that at the time of the invention, the state of the art was that methylation of integrated retroviral vectors posed serious limitations on the use of the vectors for expression of transgenes. In response, the Examiner states:

> The applicant argues that many of the references cited by Bestor and those included in Paragraph 4 of the fourth Declaration describe silencing in vitro due to methylation. This is not found persuasive for the same reasons as above. Specifically, the prior art teaches that methylation is dependent on the integration site, i.e., consistent with the teachings of Liu, Stamps, and Mathor et al. that expression level is dependent on the insertion sites. Gunzburg et al. (The EMBO Journal, 1984, 3: 1129-1135) teach that retroviral integration is random and take place either in active (i.e., the virus is expressed) or in inactive (i.e., the virus is not expressed) chromatin domains (see p. 1129, paragraph bridging columns 1 and 2, p. 1133, column 2, p. 1134, column 1). Based on these teachings, one of skill in the art would have known that the same number of integrations would result in different expression levels, depending on the insertion site. Furthermore, the prior art teaches that the expression and stability of the gene of interest directly correlates with the number of integrated retroviral vectors (see Schott et al. above). One of skill in the art would have known to look for clones comprising high numbers of integrated retroviral vectors and select the ones capable of producing high amounts of protein.

Gunzburg et al., which was published in 1984, addresses methylation of "multiple endogenous mouse mammary tumour virus (MMTV) proviral genes" that "are present at different locations in mouse inbred strains." See Abstract. Gunzburg et al. finds that the methylation patterns are location and tissue specific and that the patterns are stably inherited and appear to be conferred upon the viral DNA by the flanking mouse genomic DNA. See Abstract. The authors state that "upon integration the provirus assumes the methylation pattern of the DNA into which is integrates." p. 1129, col. 1., p. 133, col. 2. Importantly, Gunzburg et al. does not contain data or comments that address any correlation of methylation to expression of genes. Just as important, Gunzburg et al. addresses endogenous proviral sequences and not the introduction of exogenous vectors containing transgenes. These proviral sequences are endogenous to the genome and have been acquired at some point in the distant past. Gunzburg et al. has very little relevance to the present invention or to the evidence I previously submitted.

Scientific papers that are more contemporary to this invention do contain relevant evidence.

Bestor and Tycko 1996 (attached at Tab 1), identify two hypothetical roles of genomic methylation patterns. The first is a role of programmed demethylation and methylation during development. p. 363, col. 1. The second role is that cytosine methylation is part of a genome defense system which inactivates parasitic sequences such as transposable elements and proviral DNA (i.e., integrated retroviruses). p. 363, col. 1.

This second role of methylation is directly relevant to the present invention which utilizes high levels of integrated retroviral vectors. Bestor and Tycko explain this relevance:

> The host defense hypothesis requires that the silencing apparatus recognize and inactivate parasitic sequence elements. Nearly all transposition and viral integration intermediates share certain structural features, and some satellite DNA is thought to undergo amplification by extrachromosomal rolling circle replication followed by insertion of the array into the genome. Recognition and de novo methylation of CpG sites in and around features characteristic of integration reactions would insure the inactivation of the invasive element immediately upon its integration. DNA methyltransferase may have an intrinsic ability to recognize integration intermediates that are characteristic of the above integration events. The de novo sequence specificity of mammalian DNA methyltransferase is strongly dependent on alternative secondary structures in DNA; four-way junctions in cruciform structures formed by inverted-repeats in supercoiled-plasmids are especially favored targets, as are secondary structures in artificial oligonucleotide substrates. This biochemical property suggests that invasive sequences might be targeted for de novo methylation because of their presentation of alternative secondary structures during integration (Fig la).

(p. 364, col. 2, Citations omitted).

Thus, it was a concern that due to the nature of retroviral integration, the retroviral vectors would be targeted for inactivation by methylation. Increasing copy number enhances this problem. "A common characteristic of invasive sequences is their presence in multiple copies, and it has

recently become known that repeated sequences can interact so as to trigger their mutual silencing." p. 364, col. 2. Bestor and Tycko further address retroviral vectors:

> Retroviral vectors that transducer reporter genes or therapeutic agents have been observed to undergo methylation silencing after variable periods of expression in animals. Susceptability to de novo methylation and silencing has limited the usefulness of retroviral vectors in the construction of transgenic mouse lines, and it is probable that silencing phenomena may emerge as barriers to long term somatic gene therapy in humans. Successful gene transfer may require development of delivery vectors that evade the silencing response.

p. 365, col. 2.

Bestor and Tycko 1996 demonstrates why Gunzburg et al. is not relevant to the invention. Gunzburg et al. does not address the host defense mechanism at all or that fact that vectors had been shown to be actively silenced by methylation.

Garrick et al. 1998 (Tab 2) provide evidence on repeat-induced gene silencing in mammals. They used a lox/cre system to analyze the effect of copy number on transgene expression. They found that "reduction in copy number results in a marked increase in expression of the transgene and is accompanied by decreased chromatin compaction and decreased methylation at the transgene locus." p. 56, col. 1., p. 58. Again, this paper provides evidence that the state of the art was that increasing copy number leads to methylation and inactivation of transgenes. The vector construct was not a retroviral vector, but this data is directly relevant to inactivation of an introduced transgene. As Bestor and Tycko 1996 indicated, the host defense mechanism is triggered by multiple copies of invasive sequences. The transposon system used in this paper and retroviral vectors are both invasive sequences.

Cherry et al. 2000 (Tab 3) is co-authored by two of the leading scientists in the field, Dr. David Baltimore and Dr. Rudy Jaenisch. They also recognize the role of methylation in the inactivation of proviral genes. They state: "DNA methylation is thought to be a general mechanism

used by cells to silence foreign DNA and may be involved in the cell defense against transposable elements (39). DNA methylation has also been associated with the repression of gene expression and the silencing of viral control elements (2, 14, 38). Exogenously introduced retroviruses silenced in vitro and in vivo can be reactivated by treatments that result in genome wide demethylation. In addition, transcriptionally silent endogenous retroviral elements are reactivated upon loss of genomic methylation in Dnmtl knockout mice (38). Therefore, DNA methylation is thought to causally repress expression of retroviral promoters in a variety of cell types." p. 7419, col. 1-2. Both methylation-dependent and methylation-independent mechanisms exist to control retroviral gene expression." p. 7425, col. 1.

Other papers also provide evidence regarding methylation and inactivation of multiple introduced copies of exogenous genes and inaction of retroviral vectors. Mehtali et al. 1990 (Tab 4) conducted experiments that show that methylation of an introduced transgene increases with increasing copy number and that expression of the transgene decreases with increasing copy number after initially increasing. See Table 1, p. 182. The vector construct was not a retroviral vector, but this data is directly relevant to inactivation of an introduced transgene.

Niwa et al. 1983 (Tab 5) postulated that that there are two independent mechanisms that block expression from newly acquired retroviral vectors. See Abstract, p. 1105. The first mechanism operates in undifferentiated cells to block expression of M-MuLV and other exogenously acquired viral genes, such as SV40 and polyoma virus, and does not depend on DNA methylation. The second mechanism relates only to differentiated cells and represses expression of genes in which DNA is methylated. This paper further serves to demonstrate why the Examiner's reliance on Gunzburg et al. is inappropriate. Newly acquired retroviral vectors are treated by cells in a different manner from proviral sequences that have been integrated into the genome in the distant past and essentially become endogenous.

Svoboda et al. 2000 (Tab 6) examines the expression of retroviral vectors in foreign species. The vectors are subject to cell-mediated control at the transcriptional and posttranscriptional levels. Abstract, p. 181. Of main importance is cell transcriptional regulation, which can lead to proviral silencing. p. 181, col. 2. The authors note that all of the data so far point to the

important role of methylation in provirus silencing in general and that strategies for preventing methylation should contribute to more efficient gene transfer in the future. p. 186, col. 2. Again, the state of the art was that newly acquired retroviral vectors are subject to silencing by methylation. This is in direct contrast to the Examiner's conclusions based on Gunzburg et al.

Ellis and Pannell 2001 (Tab 7) also examine retrovirus silencing. They state that inclusion of appropriate regulatory elements may not be sufficient because the vectors are frequently silenced and that a better understanding of the mechanism of vector silencing is needed. p. 17, col. 1-2.

Challita and Kohn 1994 (Tab 8) provide data that shows that lack of expression following retroviral transduction is due to methylation. As stated by the authors: "Methylation of cytosine residues has been shown to be associated with suppression of gene expression and, in certain circumstances, with the silencing of viral control elements (6). The MoMuLV-LTR is completely inactive in embryonic stem and embryonic carcinoma cell lines, and the inactivity is accompanied by de novo methylation of the proviral sequences (7, 8). Moreover, methylation has been detected in association with the MoMuLV-LTR transcriptional inactivity in fibroblasts in vitro (9) and in vivo (2)." As shown by Ellis and Pannell (Tab 6), these problems still had not been solved by 2001, even when regulatory elements other than the retroviral LTR are used.

9. The references cited above establish that at the time of our invention the state of the art was that: 1) cells have a host defense mechanism that inactivates newly introduced, invading sequences such as retroviral vectors; 2) the host defense mechanism operates by methylation of the invading sequences, which causes transcriptional inactivation of the sequences; 3) transcriptional inactivation by methylation leads to reduced expression from retroviral vectors; 4) the inactivation may be triggered by structures formed during integration of the retroviral vectors; and 5) the presence of multiple repeats of an invading sequence such as a retroviral vector triggers methylation and inactivation.

10. The citations for the references cited above follow. These references are provided in the Appendix attached to this Declaration.

BESTOR TH et al, "Creation of genomic methylation patterns," Nature Genetics, 1996, Vol. 12(4) P. 363-7 (TAB 1)

GARRICK et al, "Repeat-induced gene silencing in mammals," Nature Genetics, Jan. 1998, Vol. 18(1), P. 56-9 (TAB 2)

CHERRY et al, "Retroviral Expression in Embryonic Stem Cells and Hematopoietic Stem Cells," Molecular and Cellular Biology, Oct. 2000, Vol. 20(20) P. 7419-7426 (TAB 3)

METHTALI et al, "The methylation-free status of a housekeeping transgene is lost at high copy number," Gene, 1990, Vol. 91(2), P. 179-84 (TAB 4)

NIWA et al, "Independent Mechanisms Involved in Suppression of the Moloney Leukemia Virus Genome during Differentiation of Murine Teratocarcinoma Cells," Cell, Apr. 1983, Vol. 32, P. 1105-1113 (TAB 5)

SVOBODA et al, "Retroviruses in foreign species and the problem of provirus silencing," Gene, 2000, Vol. 261, P. 181-188 (TAB 6)

ELLIS et al, "The beta-globin locus control region versus gene therapy vectors: a struggle for Expression" Clinical Genetics, Jan. 2001, Vol. 59(1) P. 17-24 (TAB 7)

CHALLITA et al, "Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo," Proc. Natl. Acad. Sci. USA, March 1994, Vol. 91, P. 2567-2571 (TAB 8)

11. I further declare that all statement made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under section 1001 of title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Dated: September 14, 2010

Dr. Gregory Bleck

TAB 1

This material may be protected by Copyright law (Title 17 U.S. Code)

Creation of genomic methylation patterns

Timothy H. Bestor[1] & Benjamin Tycko[2]

There are two biological properties of genomic methylation patterns that can be regarded as established. First, methylation of 5'-CpG-3' dinucleotides within promoters represses transcription, often to undetectable levels. Second, in most cases methylation patterns are subject to clonal inheritance. These properties suit methylation patterns for a number of biological roles, although none of the current hypotheses can be regarded as proved or disproved. One hypothesis suggests that the activity of parasitic sequence elements is repressed by selective methylation. Features of invasive sequences that might allow their identification and inactivation are discussed in terms of the genome defense hypothesis. Identification of the cues that direct *de novo* methylation may reveal the biological role (or roles) of genomic methylation patterns.

[1]*Department of Genetics and Development*
[2]*Department of Pathology, Columbia University, College of Physicians and Surgeons, 701 W. 168th Street New York, New York 10032, USA*

Correspondence should be addressed to T.H.B.

The mammalian genome is ornamented with ~3×10^7 methyl groups, all at the 5 position of cytosine (m^5C) and most at 5'-CpG-3' dinucleotides. Methylation patterns increase the information content of the genome[1] and are transmitted by clonal inheritance[2]; methylation of CpG sites within promoters represses transcription[3]. This natural modification is also dangerous: m^5C is the major endogenous mutagen (deamination results in C→T transition mutations at CpG sites, which account for about one-third of all mutations in humans)[4], and tumour suppressor genes are frequently inactivated by ectopic *de novo* methylation of promoter regions[5,6]. However, there must be benefits that yield a net selective advantage. This is shown by the retention of cytosine methylation by virtually all organisms with genomes >5×10^8 basepairs[7], and by the demonstration that perturbations of methylation patterns are lethal to mouse embryos and to differentiated cells[8,9]. While methylation patterns clearly provide some essential function, the nature of that function or functions is still enigmatic.

There are at present two salient hypothetical roles of genomic methylation patterns. The first hypothesis (put forward twenty years ago) suggests that programmed demethylation and *de novo* methylation play a direct role in gene control during development[10,11]. Methylation patterns would be established during gametogenesis or early development, and regulatory factors would mediate the removal of methyl groups from promoters to allow the expression of tissue-specific genes at the appropriate stage of differentiation. While supported by a large body of indirect and correlative evidence, a definitive test of the causality of cytosine methylation in developmental gene control has been elusive. By the same token, a developmental role for cytosine methylation has not yet been disproved. It seems likely that methylation patterns might reinforce the heritability of states of gene expression mediated by chromatin proteins analogous to the Polycomb and trithorax group of proteins from *Drosophila*[12,13].

The second hypothesis suggests that cytosine methylation is part of a genome defense system which inactivates parasitic sequences such as transposable elements and proviral DNA[4,7]. Most of these elements are in fact methylated and transcriptionally inert in the genome of mammals, flowering plants, and those fungi whose genomes contain m^5C. Treatment of cultured cells or mice with the demethylating drug 5-azacytidine can activate silent retroviruses and endogenous genes that have been silenced by ectopic *de novo* methylation of regulatory regions[14]. It is striking that *Drosophila*, whose DNA lacks m^5C, suffers far larger numbers of insertion mutations than do animals whose genomes are methylated[15]. These observations, with support from evolutionary considerations[7], strongly suggest that cytosine methylation is part of a genomic host defense system that suppresses the transcription of parasitic sequence elements. The selective advantage of such a defensive system is obvious, given that a sizable fraction of the genome represents parasitic sequences that are invisible to the immune system and which might inflict intolerable mutational or cytotoxic damage if allowed to proliferate unchecked.

These hypothetical roles of genomic methylation in development or host defense place fundamentally different requirements on the DNA methylating system. The developmental role requires that methylation patterns be established as part of the developmental program via conventional sequence recognition mediated by sequence-specific DNA methyltransferases or specificity factors that interact with the ubiquitous DNA methyltransferase. The host-defense function requires a completely different

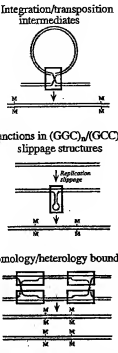


Fig. 1 Structure-dependent *de novo* methylation. Complementary sequences are depicted in identical colors, and structural features that are proposed to trigger silencing are boxed. M: 5-methylcytosine. *a*, A retrovirus integration intermediate showing the stage at which free 3' ends of the proviral DNA have been linked to 5' ends of chromosomal DNA. Most transposition and retroviral integration events utilize a common mechanism[19], and integrases and transposases share sequence or structural similarities. Recognition of integration/transposition intermediates by the silencing system could protect the genome from the proliferation of parasitic sequences. *b*, formation of three-way junctions in replication slippage structures at (GGC)/(GCC) repeats. Nearly all triplet repeat expansions at (GGC)/(GCC) sequences trigger *de novo* methylation in and around the repeats, and it is suggested that the three-way junction is the stimulus. Silencing of the *FMR1* gene in Fragile X syndrome is suggested to occur via this pathway. *c*, Detection of invasive sequences via homology/heterology boundaries during strand exchange[32]. Recognition and methylation of the boxed regions would result in reciprocal silencing of repeated sequences that reside in different sequence contexts.

source of specificity; *de novo* methylation must be cued by dynamic structural features unique to parasitic sequences in the act of invading the genome, or by events such as transient pairing interactions of repetitive elements. In other words, the developmental role requires that methylation be *directed* to specific sequences as part of the developmental programme, while the host defense role requires that methylation be *contingent* on the interaction of features characteristic of parasitic sequence elements with the DNA methylating system. Therefore it should be possible to deduce the role of methylation patterns from the nature of the biochemical events by which they were established.

Sequences that attract or repel *de novo* methylation in transfected embryonic cells have been identified[16–18], but many sequences (both foreign and endogenous) can be stably propagated in either the methylated or unmethylated state, and *de novo* methylation of such elements must therefore be dependent on features other than simple sequence recognition. The remainder of this article discusses possible mechanisms by which such sequences might be recognized and silenced.

Invasive DNA *in flagrante*

The host defense hypothesis requires that the silencing apparatus recognize and inactivate parasitic sequence elements. Nearly all transposition and viral integration intermediates share certain structural features[19], and some satellite DNA is thought to undergo amplification by extrachromosomal rolling circle replication followed by insertion of the array into the genome[20]. Recognition and *de novo* methylation of CpG sites in and around features characteristic of integration reactions would insure the inactivation of the invasive element immediately upon its integration. DNA methyltransferase may have an intrinsic ability to recognize integration intermediates that are characteristic of the above integration events. The *de novo* sequence specificity of mammalian DNA methyltransferase is strongly dependent on alternative secondary structures in DNA; four-way junctions in cruciform structures formed by inverted repeats in supercoiled plasmids are especially favored targets[21], as are secondary structures in artificial oligonucleotide substrates[22]. This biochemical property suggests that invasive sequences might be targeted for *de novo* methylation because of their presentation of alternative secondary structures during integration (Fig 1*a*).

Three-way junctions share structural features with four-way junctions, and enzymes that recognize four-way junctions (such as T4 endonuclease VII (ref. 23) and T7 endonuclease I (ref. 24)) recognize three-way junctions as well. The nature of the interaction of the methylating system with three-way junctions is of special interest because it may be involved in the aetiology of human diseases associated with GGC/GCC triplet repeat expansions. Slippage of replication intermediates can result in extrusion of a segment of GGC/GCC repeats, which form stable three-way junctions despite a lack of perfect complementarity in the extruded segment[22,23](Fig. 1*b*). The preference of DNA methyltransferase for junctions may result in *de novo* methylation of sequences in slippage structures. Diseases that result from GGC/GCC triplet repeat expansions may be thought of as autoimmune disorders, in which a host defense system attacks an innocuous alteration of an endogenous gene because of its incidental similarity to a parasitic sequence element in the act of integration.

Cytosine methylation and pairing interactions

A common characteristic of invasive sequences is their presence in multiple copies, and it has recently become known that repeated sequences can interact so as to trigger their mutual silencing. Fungi and flowering plants have diverse and highly effective means of silencing repeated sequences, and cytosine methylation is associated with silencing in nearly all cases. RIP (repeat-induced point mutation) imposes methylation, silencing, and large numbers of C→T transition mutations on repeated sequences during the sexual phase of the fungus *Neurospora crassa*[26], while MIP (methylation induced premeiotically) inactivates and methylates repeats during the sexual cycle of the fungus *Ascobolus immersus*[27]. Flowering plants can also silence and methylate repeated sequences via RIGS (repeat-induced gene silencing)[28,29]; the efficiency is such that multicopy transgenes frequently cause mutual silencing of themselves and of homologous resident sequences. The process can be remarkably efficient; unlinked transgene sequences as small as 300 bp can identify and inactivate each other in a genome of $>10^9$ bp.

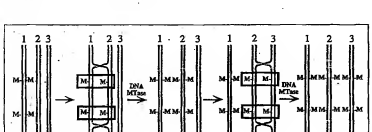


Fig. 2 Unidirectional transfer of epigenetic information via paired hemimethylated intermediates. Strand exchange between methylated and unmethylated repeated sequences presents hemimethylated sites, which are the preferred substrate of DNA methyltransferase; note that the maintenance activity of DNA methyltransferase leads to *de novo* methylation under these conditions. A common methylation pattern could propagate through an array or network of repeats by this mechanism, and methylation patterns could be transferred between alleles at loci that bear allele-specific methylation patterns.

The silenced state can persist even after the repeated sequences have been separated by segregation in sexual crosses[28]. Transgene silencing of this type is emerging as a barrier to the improvement of commercially-important plant species[30,31]. It is difficult to imagine a selective advantage for RIP, MIP, or RIGS other than host defense against the proliferation of parasitic sequence elements.

Transient pairing interactions are likely to be involved in the imposition of silencing and methylation on repeats; two mechanisms can be envisioned. First, strand exchange between repeats at non-allelic positions will present an abrupt loss of homology at the junction with flanking sequences (Fig. 1*c*), and silencing may be provoked by factors that recognize characteristic structural features at the homology/heterology boundary. This idea was developed by Signer and colleagues[32] to explain copy number-dependent transgene silencing in *Arabidopsis*. Second, methylation might propagate through a network of repeated sequences once one or more copies are methylated (Fig. 2). Strand exchange between methylated and unmethylated repeated sequences creates hemimethylated intermediates, which as mentioned previously, are strongly preferred substrates of DNA methyltransferase[12]; the preference for hemimethylated substrates normally contributes to maintenance methylation, but under these conditions *de novo* methylation is the result. Strand exchange reactions (which are thought to occur frequently as part of the double strand break repair pathway[33]) (Fig. 2) could cause a common methylation pattern to propagate through a network of repeats[28,30], and the probability of pairing with a methylated repeat will increase in proportion to the number of methylated repeats. A large number of methylated resident repeats will therefore increase the probability that a new, unmethylated copy will be methylated soon after its insertion into the genome. A large burden of parasitic sequences may have the paradoxical effect of repressing their own activity and that of homologous invasive sequences. Natural selection may actually favor the retention of large numbers of inactive parasitic sequences for this reason.

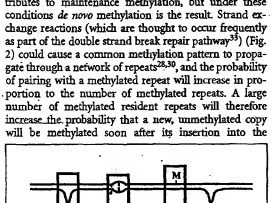


Fig. 3 Vulnerability of heteroduplex recombination intermediates to *de novo* methylation. Crossing-over during meiosis presents DNA methyltransferase with highly preferred targets: four-way junctions, mismatches and hemimethylated sites. Dysregulated *de novo* methylation appears to be prevented by sharp down-regulation of DNA methyltransferase during the pachytene stage of meiosis; the mechanism involves the production of a larger, non-translated DNA methyltransferase mRNA[48].

Many endogenous genes are present in multiple copies which escape repeat-dependent gene silencing. In *Neurospora*, which seems to be especially aggressive in its response towards repeated sequences, the efficiency of RIP is greater when the repeats are in close proximity[26]. This argues for special protective mechanisms that shield repeated cellular genes from repeat-induced silencing. A complete intolerance of repeated sequences would also put severe constraints on the evolution of new functions, which depends on the duplication and divergence of existing genes. However, the factors that control sensitivity to repeat-dependent silencing are poorly understood[34].

There is no clear experimental evidence that repeat-dependent silencing or methylation operates in mammals, as is the case in fungi and plants. However, no objective search for repeat-dependent silencing in mammals seems to have been conducted. Circumstantial evidence suggests that it may occur[35]. Much of the m^5C in the mammalian genome is found in repetitive DNA, and most of the repeated and potentially transposable elements in the genome are methylated and transcriptionally inert through most of development[36,37]. Retroviral vectors that transduce reporter genes or therapeutic agents have been observed to undergo methylation silencing after variable periods of expression in animals[38]. Susceptibility to *de novo* methylation and silencing has limited the usefulness of retroviral vectors in the construction of transgenic mouse lines, and it is probable that silencing phenomena may emerge as barriers to long-term somatic gene therapy in humans. Successful gene transfer may require the development of delivery vectors that evade the silencing response. It is also possible that the development of therapeutic agents that perform selective methylation of deleterious sequence elements (such as HIV-1 proviral DNA) will activate an existing host defense system and therefore serve as a sort of nuclear vaccine.

Maintenance of allele-specific methylation patterns

Mammalian genes whose expression is dependent on the sex of the contributing parent are said to be imprinted; such genes bear allele-specific methylation patterns that are necessary for maintenance of the imprinted state[9]. Short repetitive sequences are characteristic of imprinted genes[39,40]; these short repeats may form alternative secondary structures that trigger *de novo* methylation in the germ line or during early postfertilization development. A parent-of-origin effect on gene expression results if there

are differences in the types of DNA or chromatin structures formed, or if the response to the structures differs, in the male and female germlines. In this way a nuclear host defense system may give rise to allele-specific methylation patterns and to the genetic phenomenon of imprinting[41], and could also account for documented gamete-of-origin effects on methylation of repetitive DNA[37]. It should be noted that several mutations that result from insertion of retroviruses of the intracisternal A particle (IAP) type confer imprinted behavior on genes that are not imprinted in the wild type[42,43].

Alleles at imprinted loci are asymmetrically methylated, and the mechanism shown in Fig. 2 will tend to convert asymmetrical allelic methylation patterns towards the more heavily methylated pattern. Certain pathological human conditions show abnormalities in the functional imprinting of particular chromosomal regions which might arise via this type of interchromosomal transfer of epigenetic information. For example, Wilms' tumours frequently show conversion to a uniparental (bipaternal) methylation and expression pattern at imprinted loci in the *H19/IGF2* region on chromosome 11p15.5, which could result from the local transfer of methylation patterns from the paternal chromosome to the less heavily-methylated maternal chromosome[44]. It is important to note, however, that a conversion to a symmetrical uniparental allelic methylation pattern at imprinted loci could also come about by an alternative mechanism that involves deletion or mutations of a *cis*-acting DNA element (an 'imprinting centre'). This has been implicated in certain kindreds that show disruption of imprinting patterns within the chromosome 15 Prader Willi/Angelman syndrome region[45].

Vulnerability of meiotic recombination intermediates

Allelic heteroduplex recombination intermediates present several features that should make them very vulnerable to *de novo* methylation (Fig. 3). First, allelic methylation differences create hemimethylated sites in the heteroduplexes, which provoke *de novo* methylation via the maintenance activity of DNA methyltransferase (see Fig. 2). Second, Holliday structures and four-way junctions are necessarily present; as described earlier, these structures appear to be favored targets of DNA methyltransferase. Third, mismatches in the vicinity of CpG dinucleotides greatly favor *de novo* methylation, presumably by lowering the energetic barrier associated with eversion of the target cytosine during the transmethylation reaction[46,47]. Presentation of these vulnerable sites might create a "methylation ratchet" in which methylation levels increase in an unregulated fashion with each meiotic cycle.

Examination of purified germ cells from male mice showed that both DNA methyltransferase protein and the 5.2-kb DNA methyltransferase mRNA found in all proliferating cell types was present in all germ cell fractions, except for pachytene spermatocytes. These spermatocytes showed an absence of DNA methyltransferase protein and contained a 6.2-kb RNA that was not associated with polyribosomes. It is the pachytene stage of meiosis I where most crossing-over occurs. The ubiquitous 5.2-kb DNA methyltransferase mRNA and DNA methyltransferase protein reappeared at the conclusion of the crossing-over phase of meiosis[48]. These findings suggest that meiotic recombination intermediates are protected from *de novo* methylation through down-regulation of DNA methyltransferase via a novel post-transcriptional mechanism that involves the production of a larger, non-translated RNA transcript.

The function of cytosine methylation

The fact that cytosine methylation can increase the information content of DNA has tempted many to attribute diverse roles to methylation patterns. Cytosine methylation has been proposed to reduce the effective size of the genome by masking non-regulatory regions in large-genome eukaryotes[7,49], and central roles in DNA repair[22,50] and replication[21] have also been mooted. None of the hypothetical functions of cytosine methylation (and this includes the developmental and host-defense functions) has the support of compelling experimental evidence, and all, some, or none of the hypotheses may be valid. At this time it seems that the true function of cytosine methylation will be understood only when we learn how the cell selects specific sequences for covalent modification.

Acknowledgements

E. Selker and M. Turker made helpful comments on the manuscript. Supported by NIH grants GM00616 and CA60610 (to T. H. B.) and NIH grant CA60765 and American Cancer Society grant JFRA482 (to B.T.).

Received 27 November 1995; accepted 3 January 1996.

1. Reinisch, K. M., Chen, L., Verdine , G.L. & Lipscomb, W.N. (1995) The crystal structure of *HaeIII* methyltransferase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell* 82, 143–153 (1995).
2. Pfeifer, G., Steigerwald, S., Hansen, R.S. Gartler, S.M. & Riggs, A.D. (1990) Polymerase chain reaction aided genomic sequencing of an X chromosome linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl. Acad. Sci. USA* 87, 8252–8256 (1990).
3. Iguchi-Ariga, S.M.M. & Schaffner, W. CpG methylation of the cAMP responsive enhancer/promoter TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* 3, 612–619 (1989).
4. Bestor, T.H. & Coxon, A. The pros and cons of cytosine methylation *Curr. Biol.* 3, 384–386 (1993).
5. Yoshiura, K. *et al.* Silencing of the E-cadherin invasion-suppressor gene by CpG methylation in human carcinomas. *Proc. Natl. Acad. Sci. USA* 92, 7416–7419 (1995).
6. Herman, J. *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. USA* 91, 9700–9704 (1994).
7. Bestor, T.H. DNA methylation: evolution of a bacterial immune function into a regulator of genome structure and activity in higher eukaryotes. *Phil. Trans. Royal Soc. Lond.* B 326, 179–187 (1989).
8. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of mouse DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926 (1992).
9. Li, E., Beard, C., Forster, A.C., Bestor, T.H. & Jaenisch, R. DNA methylation, genomic imprinting, and mammalian development. *Cold Spring Harb. Symp. Quant. Biol.* LVIII, 297–305 (1993).
10. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* 187, 226–232 (1975).
11. Riggs, A.D. X-inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* 14, 9–11 (1975).
12. Bestor, T.H. & Verdine, G.L. DNA methyltransferases. *Curr. Opin. Cell Biol.* 6, 380–389 (1994).
13. Bestor, T.H. The biochemical basis of allele-specific gene expression in genomic imprinting and X inactivation. In *Genomic imprinting: Causes and consequences.* (eds Ohlsson, R., Hall, K. & Ritzen, M.) (Cambridge University Press, Cambridge, 1995).
14. Jaenisch, R., Schnieke, A. & Harbers, K. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. *Proc. Natl. Acad. Sci. USA* 82, 1451–1455 (1985).
15. Ashburner, M. *Drosophila: A Laboratory Handbook.* 103 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1992).
16. Mummenani, P., Bishop, R. & Tucker, M. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. *J. Biol. Chem.* 268, 552–558 (1993).
17. Macleod, D., Charlton, J., Mullins, J. & Bird, A. Sp1 binding sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev.* 8, 2282–2292 (1994).
18. Brandeis, M. *et al.* Sp1 elements protect a CpG island from *de novo* methylation. *Nature* 371, 435–438 (1994).
19. Craig, N.L. Unity in transposition reactions. *Science* 270, 253–255 (1995).
20. Charlesworth, B., Sniegowski, P. & Stephan, W. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371, 215–220 (1994).
21. Bestor, T.H. Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. *Nucl. Acids Res.* 15, 3635–3643 (1987).
22. Chen, X. *et al.* Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. USA* 92, 5199–5203 (1995).
23. Iliakis, F. & Kemper, B. Endonuclease VII resolves Y-junctions in branched DNA *in vitro*. *EMBO J.* 5, 181–190 (1986).
24. Lu, M., Guo, Q., Studier, W.F. & Kallenbach, N.F. Resolution of branched DNA substrates by T7 endonuclease I and its inhibition. *J. Biol. Chem.* 266, 2531–2536 (1991).
25. Usdin, K. & Woodford, K.J. CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro. *Nucl. Acids Res.* 23, 4202–4209 (1995).
26. Selker, E.U. Premiotic instability of repeated sequences in *Neurospora crassa*. *Annu. Rev. Genet.* 24, 579–613. (1990).
27. Goyon, C., Noguiès, T.L.V. & Faugeron, G. Perpetuation of cytosine methylation in *Ascobolus immersus* implies a novel type of maintenance methylase. *J. Mol. Biol.* 240, 42–51 (1994).
28. Matzke, M.A. & Matzke, A.J.M. Homology-dependent gene silencing in transgenic plants: what does it really tell us? *Trends Genet.* 11, 1–3 (1995).
29. Matzke, A.J.M., Neuhuber, F., Park, Y.D., Ambros, P.F. & Matzke, M.A. Homology-dependent gene silencing in transgenic plants: epistatic silencing loci contain multiple copies of methylated transgenes *Mol. Gen. Genet.* 244, 219–229 (1994).
30. Flavell, R.B. Inactivation of gene expression in plants as a consequence of specific sequence duplication. *Proc. Natl. Acad. Sci. USA* 91, 3490–3496 (1994).
31. Finnegan, J. & McElroy, D. Transgene inactivation: plants fight back! *Biotechnol.* 12, 883–888. (1994).
32. Assaad, F.F., Tucker, K.L. & Signer, E.R. Epigenetic repeat-induced gene silencing (PIGS) in *Arabidopsis*. *Plant Mol. Biol.* 22, 1067–1085 (1993).
33. Engels, W.R., Preston, C.R. & Johnson-Schitz, D.M. Long-range cis preference in DNA homology search over the length of the Drosophila chromosome. *Science* 263, 1623–1625 (1994).
34. Bender, J. & Fink, G. Epigenetic control of endogenous gene family is revealed by a novel blue fluorescent mutant of *Arabidopsis*. *Cell* 83, 725–734 (1995)
35. Kricker, M.C., Drake, J.W. & Radman, M. Duplication-targeted DNA methylation and mutagenesis in the evolution of eukaryotic chromosomes. *Proc. Natl. Acad. Sci. USA* 89, 1075–1079 (1992).
36. Chesnokov, I.N. & Schmid, C.W. Specific Alu binding protein from human sperm chromatin prevents DNA methylation. *J. Biol. Chem.* 270,18539–18542 (1995)
37. Rubin, C.M., VandeVoort, C.A., Tipitiz, R.L. & Schmid, C.W. Alu repeated DNAs are differentially methylated in primate germ cells. *Nucl. Acids Res.* 22, 5121–5127 (1994).
38. Chaillta, P.-A. & Kohn, D.B. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*. *Proc. Natl. Acad. Sci. USA* 91, 2567–2571 (1994).
39. Neumann, B., Kubicka, P. & Barlow, D.P. Characteristics of imprinted genes. *Nature Genet.* 9, 451–452 (1994).
40. Hatada, I. & Mukai, T. Genomic imprinting of *p57KIP2*, a cyclin-dependent kinase inhibitor in mouse. *Nature Genet.* 11, 204–205 (1995).
41. Barlow, D.P. Methylation and imprinting: from host defense to gene regulation? *Science* 260, 309–310 (1993).
42. Michaud, E. *et al.* Differential expression of a new dominant agouti allele (A^{iapy}) is correlated with methylation state and is influenced by parental lineage. *Genes Dev.* 8, 1463–1472 (1994).
43. Duhl, D., Vrieling, H., Miller, K., Wolff, G. & Barsh, G. Neomorphic *agouti* mutations in obese yellow mice. *Nature Genet.* 7, 220–226 (1994).
44. Moulton, T. *et al.* Epigenetic lesions at the *H19* locus in Wilms' tumour patients. *Nature Genet.* 7, 440–447 (1994).
45. Bulling, K. *et al.* Inherited microdeletions in the Angelman and Prader-Will syndromes define an imprinting centre on human chromosome 15. *Nature Genet.* 9, 395–401 (1995).
46. Yang, A.S., Shen, J.-C., Zingg, J.-M., Mi, S. & Jones, P.A. HhaI and HpaII DNA methyltransferases bind DNA mismatches, methylate uracil and block DNA repair. *Nucl. Acids Res.* 23, 1390–1397 (1995).
47. Klimasauskas, S. & Roberts, R.J., M.HhaI binds tightly to substrates containing mismatches at the target base. *Nucl. Acids Res.* 23, 1388–1395 (1995).
48. Jue, K., Bestor, T.H. & Trasler, J. Regulated synthesis and localization of DNA methyltransferase during spermatogenesis. *Biol. Reprod.* 53, 561–569 (1995).
49. Bird, A.P. Gene number, noise reduction and biological complexity. *Trends Genet.* 11, 94–100 (1995).
50. Cleaver, J. It was a very good year for DNA repair. *Cell* 76, 1–4 (1994)
51. Taylor, J.H. Origins of replication and gene regulation. *Mol. Cell. Biochem.* 61, 99–109 (1984).

TAB 2

This material may be protected by Copyright law (Title 17 U.S. Code)

Repeat-induced gene silencing in mammals

David Garrick[1], Steven Fiering[2,3], David I.K. Martin[2] & Emma Whitelaw[1]

In both plants[1-3] and *Drosophila melanogaster*[4,5], expression from a transgenic locus may be silenced when repeated transgene copies are arranged as a concatameric array. This repeat-induced gene silencing is frequently manifested as a decrease in the proportion of cells that express the transgene, resulting in a variegated pattern of expression. There is also some indication that, in transgenic mammals, the number of transgene copies within an array can exert a repressive influence on expression, with several mouse studies reporting a decrease in the level of expression per copy as copy number increases[6-8]. However, because these studies compare different sites of transgene integration as well as arrays with different numbers of copies, the expression levels observed may be subject to varying position effects as well as the influence of the multicopy array. Here we describe use of the lox/Cre system of site-specific recombination to generate transgenic mouse lines in which different numbers of a transgene are present at the same chromosomal location, thereby eliminating the contribution of position effects and allowing analysis of the effect of copy number alone on transgene silencing. Reduction in copy number results in a marked increase in expression of the transgene and is accompanied by decreased chromatin compaction and decreased methylation at the transgene locus. These findings establish that the presence of multiple homologous copies of a transgene within a concatameric array can have a repressive effect upon gene expression in mammalian systems.

The αPElox construct used to generate transgenic mouse lines is shown in Fig. 1*a*. In this transgene, expression of the *lacZ* reporter gene is driven by the human α-globin promoter and the αHS-40 enhancer-like element of the α-like globin locus[9]. As expression of the *lacZ* reporter can be analysed in single erythroid cells, use of this transgene allows the detection of variegated patterns of expression[10]. The transgene also contains a single copy of the 34-bp recognition site (*loxP*) of the Cre recombinase of bacteriophage P1 (ref. 11). Cre-mediated recombination between *loxP* sites in individual αPElox transgenes that form part of a concatameric array will reduce the transgene copy number without altering the site of integration[12] (Fig. 1*b*). Although transgenes within a multicopy array in mice are usually present in a tandem (head–tail) orientation, inverted repeats (head–head or tail–tail) do occur[13]. Because Cre recombination between inverted *lacZ* sites causes the inversion rather than excision of the intervening DNA, only recombination between sites in like-oriented transgene monomers will reduce the copy number of the array.

We generated founder transgenic mice bearing the αPElox construct, and from them we established hemizygous transgenic lines. Southern blotting of tail DNA revealed that two of these transgenic lines (αPElox1 and αPElox7) contain more than 100 copies of the transgene (Fig. 2), and these lines were selected for Cre-mediated reduction in transgene copy number. For each of these parent lines, we collected fertilized oocytes from wild-type female mice mated with G_1 hemizygous transgenic males. Oocytes were micro-injected with a circular Cre expression vector (pCAGGS-Cre) and then transferred into foster mothers. Transient expression of the recombinase from the unincorporated pCAGGS-Cre plasmid has been shown to catalyze efficient site-specific recombination at *loxP* sites within the mouse genome before the morular stage of development[14]. Live-born progeny that showed a reduced transgene copy number as determined by initial Southern blotting were mated with wild-type animals to establish transgenic lines hemizygous for the modified locus. Cre recombination within the αPElox1 parent line (more than 100 copies) gave rise to two distinct reduced-copy progeny lines: 1.cre/a, which contains five copies of the transgene, and 1.cre/b, in which the array has been reduced to a single copy (Fig. 2). Southern-blot analysis indicated that the single copy remaining in the 1.cre/b line contains a rearrangement/deletion (data not shown). The αPElox7 parent line (more than 100 copies) gave rise to a single derivative line (7.cre/a) bearing one copy of the transgene.

For parental and reduced-copy progeny lines, transgene expression was analysed in 12.5-dpc embryos by staining of whole primitive erythrocytes with X-gal. We previously showed that all cells containing β-galactosidase activity can be detected by light microscopy after staining under these conditions[10,15]. For both of the high-copy parent lines, a heavily variegated pattern of transgene expression was observed, with less than 1% of

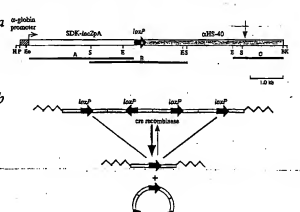


Fig. 1 Strategy for Cre-mediated reduction of transgene copy number. *a*, The αPElox transgene construct contains nucleotides −573 to +36 of the human α-globin promoter upstream of the SDK-*lacZ*pA cassette used previously[30] and a 4.2-kb fragment containing the αHS-40 DNaseI hypersensitive site (vertical arrow). An oligonucleotide containing a single copy of the 34-bp *loxP* site was inserted between the SDK-*lacZ*pA cassette and the αHS-40 fragment. Black lines indicate probes used in this study. H, *Hin*dIII; P, *Pst*I; Eo, *Eco*0109 I; S, *Sac*I; E, *Eco*RI; B, *Bam*HI; K, *Kpn*I. *b*, When multiple copies of the αPElox transgene are situated within a concatameric array in the mouse genome, Cre-mediated recombination between like-oriented *loxP* sites within the array removes the intervening DNA as a circular episome, leaving a reduced-copy array at the same chromosomal location. The forward (excision) reaction is heavily favoured over the reverse (integration) reaction, which requires recombination between *loxP* sites on separate molecules.

[1]*Department of Biochemistry, University of Sydney, New South Wales 2006, Australia.* [2]*Fred Hutchinson Cancer Research Center, Seattle, Washington, USA.* [3]*Current address: Department of Microbiology, Dartmouth Medical Center, Lebanon, New Hampshire, USA. Correspondence should be addressed to E.W. e-mail: e.whitelaw@biochem.usyd.edu.au*

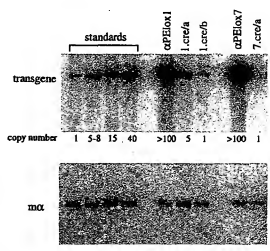


Fig. 2 Copy-number determination for αPElox parental and Cre-modified progeny transgenic lines. DNA was obtained from tail biopsies of three-week-old G$_1$ hemizygous mice for each of the lines shown, as well as from lines bearing known copy numbers of a transgene that contains the same *lacZ*–αHS-40 cassette. After digestion with *Sac*I, DNA was resolved on a 0.8% agarose gel and transferred to a nylon membrane. The membrane was hybridized with fragment B (Fig. 1a), and the copy number was determined by quantifying the intensity of the resulting signal and comparing it with those of the standards. To ensure equivalent DNA loadings in each lane, the membrane was stripped and re-hybridized with a probe for the murine α-globin gene (mα). In a separate Southern blot performed on tail DNA from each of these lines, hybridization with a probe specific for the 3′ end of the transgene (fragment C, Fig. 1a) revealed that the junction fragments between insert and mouse DNA generated with the restriction enzymes *Eco*RI or *Eco*RV for each of the reduced-copy progeny lines were the same as those observed in the respective parent arrays, indicating that the genomic location of the array had not been altered during Cre modification (data not shown).

erythroid cells containing an active transgene locus (Fig. 3, Table 1). Similar findings of variegated expression patterns in high-copy transgenic lines bearing a construct (αPE) that differs from αPElox only in the absence of the single *loxP* element[16] suggest that the presence of the 34-bp *loxP* oligonucleotide has no effect on transgene expression. When the number of transgene copies within the αPElox1 array was reduced to five by Cre-mediated recombination (1.cre/a), there was a large increase (more than 1,000-fold) in the percentage of primitive erythroid cells expressing the transgene. Similarly, Cre-mediated reduction of the αPElox7 parent array to one copy (7.cre/a) also suppressed the variegated expression that was observed in the parent line, with a 180-fold increase in the size of the expressing population (Fig. 3, Table 1). In these two independent transgenic lines, the presence of high-copy multimeric arrays is therefore associated with a silencing of transgene expression, which is observed as a decrease in the percentage of cells containing an active transgene locus. A decrease in the number of copies within each array correlates with a suppression of variegation—that is, an increase in the size of the expressing population. No expression of the transgene was observed in the line 1.cre/b, in which the single transgene copy was rearranged (data not shown).

In plants, repeat-induced gene silencing has been observed at both the transcriptional[17–19] and post-transcriptional[20,21] stages of gene expression. To determine whether the copy-number–dependent silencing of transgene expression in mouse erythroid cells occurs at transcription or involves a post-transcriptional modification, we performed run-on analysis in nuclei of 12.5-dpc erythrocytes from αPElox1 and its reduced-copy derivative, 1.cre/a (Fig. 4*a*). Run-on transcripts from the *lacZ* reporter gene were present in 1.cre/a primitive erythrocytes but were not detectable in erythroid cells of the high-copy parent line, indicating that the silencing of expression from multi-copy transgene arrays occurs at the level of transcription. The methylation status and local chromatin structure of the transgene locus in these two lines were also compared. Although transgenes present in 12.5-dpc erythrocytes of the reduced-copy line 1.cre/a appear unmethylated, the high-copy αPElox1 transgene locus at the same genomic location was found to be heavily methylated (Fig. 4*b*). To analyse chromatin structure, we performed endonuclease protec-

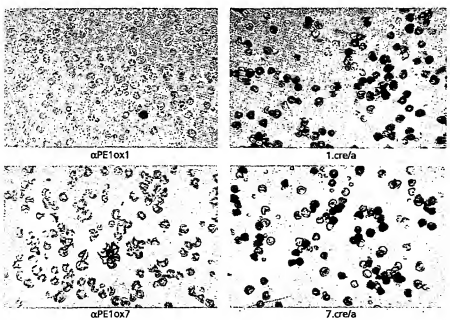


Fig. 3 X-gal staining of primitive erythrocytes from the αPElox transgenic lines and their reduced-copy progeny. Individual 12.5-dpc embryos from the transgenic lines shown were bled into PBS and whole blood cells were stained with X-gal before visualization under light microscopy. For αPElox1, many fields of view had to be scanned to detect a single blue cell.

Table 1 • Copy number and erythroid expression

Transgenic line	Copy number	Percentage of 12.5-dpc erythroid cells expressing *lacZ*
αPElox1	>100	0–0.01 (n=11)
1.cre/a	5	65±9 (n=11)
αPElox7	>100	0.3±0.02 (n=4)
7.cre/a	1	54±9 (n=3)

Transgene copy numbers were determined by Southern analysis (Fig. 2). The percentage of 12.5-dpc erythroid cells expressing *lacZ* was determined by staining cells as described (Fig.3) and scoring a minimum of 200 cells for a detectable blue colour. Expression data are presented as the mean ± one standard deviation—except for αPElox1, where a range is given. n, number of individual transgenic embryos assayed for each line.

tion assays[22] on nuclei isolated from 12.5-dpc primitive erythroid cells from these two lines. In their native chromatin configuration, transgenes in the high-copy array (αPElox1) were more resistant to endonuclease digestion at a site within the transgene promoter than transgenes that were at the same genomic location but within a lower-copy array (1.cre/a; Fig. 4c). The transcriptional silencing that occurs at high-copy arrays is therefore associated both with hypermethylation of transgenes in the locus and with the adoption of a repressive local chromatin configuration.

These results establish that the presence of multiple repeats within a high-copy array can directly repress transgene expression in a mammalian system. Because transgenic animals frequently contain high-copy concatameric arrays of a monomeric unit, this observation has important implications for the meaningful interpretation of transgenic experiments in mammals. The influence of copy number on silencing is likely to be most prominent in lines such as those studied here, in which the array is very large, but repeat-induced silencing may often be responsible for poor transgene expression. In addition to copy number, as shown here, the factors influencing transgene silencing in mice include integration site, the lineage in which expression occurs and the *cis*-acting control elements within the transgene[23]. Transgene constructs for which copy-number–dependent expression has been reported[24,25] may contain genetic elements that function to insulate individual monomers and prevent silencing. Without such elements, the lox/Cre system of site-specific recombination could be used to activate expression in high-copy lines.

The hypermethylation of transgenes within the inactive, high-copy array noted here has also been reported in other cases of repeat-induced silencing[1,3,26,27], although it is unclear in our system whether methylation precedes or is a consequence of the observed chromatin restructuring. The occurrence of copy-number–dependent transgene inactivation in *Drosophila*, in which DNA methylation has not been detected[28], suggests that methylation is not necessary for the chromatin remodelling associated with repeat-induced silencing and that hypermethylation of the αPElox transgene in mice may be a secondary modification that occurs at already inactive, high-copy arrays. This hypothesis is consistent with our previous finding that the variegated silencing of globin/*lacZ* transgenes in mice correlates with an inactive chromatin structure but not hypermethylation at the transgene locus[29].

Although a molecular mechanism for repeat-induced silencing of multimeric arrays is yet to be fully elucidated, the correlation between the silenced state and the adoption of a less accessible

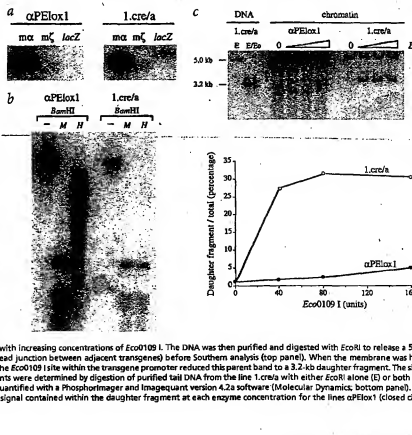


Fig. 4 Transcription, methylation and chromatin accessibility of the transgene. **a**, Nuclear run-on analysis of *lacZ* transcription in αPElox1 and 1.cre/a. Nuclei were prepared from primitive erythroid cells collected from 12.5-dpc transgenic embryos of the lines αPElox1 and 1.cre/a. Run-on transcripts were synthesized and hybridized to membranes containing the following DNA fragments: the mouse α-globin gene from +17 to +956 (mα), the mouse ζ-globin gene from +150 to +969 (mζ) and a fragment of the *lacZ* gene, which includes nucleotides from +39 to +1164 relative to the start site of the αPElox transcript (*lacZ*). **b**, Analysis of methylation status of the transgene locus in αPElox1 and 1.cre/a. DNA from 12.5-dpc erythrocytes of the lines αPElox1 and 1.cre/a was digested with either *Bam*HI alone or *Bam*HI together with either *Msp*I (M) or *Hpa*II (H), which are insensitive and sensitive, respectively, to methylation at the central CpG dinucleotide of their common recognition sequence, 5′-CCGG. The resulting fragments were analysed by Southern blotting and hybridization with fragment A (Fig. 1a). **c**, Endonuclease protection assay of the transgene locus in αPElox1 and 1.cre/a. Nuclei prepared from primitive erythroid cells of the lines αPElox1 or 1.cre/a were digested with increasing concentrations of *Eco*0109 I. The DNA was then purified and digested with *Eco*RI to release a 5.0-kb parent fragment (composed of the tail–head junction between adjacent transgenes) before Southern analysis (top panel). When the membrane was hybridized with fragment A (Fig. 1a), cleavage at the *Eco*0109 I site within the transgene promoter reduced this parent band to a 3.2-kb daughter fragment. The sizes of the expected parent and daughter fragments were determined by digestion of purified tail DNA from the line 1.cre/a with either *Eco*RI alone (E) or both *Eco*RI and *Eco*0109 I (E/Eo). The membrane was quantified with a PhosphorImager and ImageQuant version 4.2a software (Molecular Dynamics; bottom panel). The vertical axis shows the percentage of total signal contained within the daughter fragment at each enzyme concentration for the lines αPElox1 (closed circles) and 1.cre/a (open circles).

chromatin configuration observed both here and in *Arabidopsis*[19] is consistent with a model in which homologous pairing between monomers within the array induces heterochromatinization at the transgene locus. Support for a model of heterochromatin formation is derived from the suppression of copy-number–dependent silencing of pigment genes in the *Drosophila* eye by mutations within genes encoding known structural components of heterochromatin[4]. It remains to be determined whether the repeat-induced modification of chromatin structure in mice is dependent on the proximity of the locus to nearby blocks of constitutive heterochromatin, as was the case for a *brown* eye pigment transgene in *Drosophila*[2], or whether high-copy arrays autonomously form inactive chromatin structures irrespective of their position. The arrangement of endogenous loci such as the rRNA, tRNA and histone genes as high-copy concatamers of a repeated unit suggests that multi-copy arrays need not always be subject to repeat-induced silencing. An investigation of how silencing is prevented at these endogenous loci may prove useful in maintaining activity at multi-copy arrays of foreign genetic elements.

Methods

Transgenic mice. The αPElox transgene was constructed by insertion of an oligonucleotide containing a single copy of the 34 bp *loxP* site (5′-ATAACTTCGTATAATGTATGCTA TACGAAGTTAT-3′) between the SDK–*lacZpA* cassette and the 4.2-kb αHS-40 fragment of the previously described construct αPE[14]. The integrity of the *loxP* site was confirmed by dideoxy sequencing. The αPElox fragment for micro-injection was excised from plasmid vector sequences by digestion with *Hin*dIII and *Kpn*I and purified by agarose-gel electrophoresis. Transgenic mouse lines bearing the αPElox construct were generated by standard micro-injection techniques in the outbred Pathology Oxford (P.O.) mouse strain. For Cre recombination, fertilized oocytes were collected from wild-type P.O. females mated with hemizygous G_1 transgenic male mice and micro-injected into either pronuclei with the circular Cre-expression vector pCAGGS-Cre[14] at 5 ng/μl before transfer into pseudo-pregnant foster mothers. Live-born progeny that showed a reduced transgene copy number as determined by initial Southern blotting were mated with wild-type animals to establish transgenic lines hemizygous for the modified locus.

Histology. After individual 12.5-dpc embryos had been bled into PBS, whole blood cells were gently pelletted and then fixed in 0.25% glutaraldehyde before staining with 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside (X-gal) at 37 °C for 24 h as described previously[20].

Nuclear run-on and endonuclease protection assays. To prepare membranes for nuclear run-on analysis, probe fragments were excised from plasmid vector by digestion with the appropriate restriction enzymes and purified by agarose-gel electrophoresis. Probe fragments used were an *Hin*fI–*Hin*fI fragment containing +17 to +956 relative to the transcription start site of the mouse ζ-globin gene; an *Xba*I–*Pst*I fragment containing +150 to +969 of the mouse ζ-globin gene; and a *Pst*I–*Eco*RV fragment of pSDK–*lacZpA*[20], which includes +39 to +1164 relative to the start site of the *lacZ* transcript produced by αPElox. Purified fragment DNA (0.2 μg of each fragment) was then electrophoresed on 1.2% agarose gel and transferred to a nitrocellulose membrane by Southern blotting. Nuclei were purified as described[5] from 1.5×10^7 primitive erythroid cells collected from 12.5-dpc transgenic embryos. Run-on transcripts were synthesized from isolated nuclei and hybridized to membranes as described previously[21]. For endonuclease protection assays, nuclei prepared as described above were separated into four aliquots of 250 μl and digested with 0, 40, 80 or 160 U of *Eco*0109 I at 37 °C for 90 min. After proteinase-K digestion and phenol-chloroform extraction, purified DNA was digested with *Eco*RI and Southern analysis performed.

Acknowledgements

We thank P. Vassalli for the gift of the pCAGGS-Cre plasmid, H. Sutherland for the gift of the αPE plasmid, W. Magin for technical support and M. Groudine for support and helpful discussions. This work was supported by a grant from the National Health and Medical Research Council of Australia (to E.W.), an Australian Postgraduate Award (to D.G.), an NIH grant (to D.I.K.M.) and a grant from the Burroughs-Wellcome Fund to S.F. D.I.K.M. is a scholar of the Leukemia Society of America.

Received 6 August; accepted 10 October, 1997.

1. Linn, F., Heidmann, I., Saedler, H. & Meyer, P. Epigenetic changes in the expression of the maize A1 gene in *Petunia hybrida*: role of numbers of integrated gene copies and state of methylation. *Mol. Gen. Genet.* **222**, 329–336 (1990).
2. Scheid, O.M., Paszkowski, J. & Potrykus, I. Reversible inactivation of a transgene in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **228**, 104–112 (1991).
3. Assaad, F.F., Tucker, K.L. & Signer, E.R. Epigenetic repeat-induced gene silencing (RIGS) in *Arabidopsis*. *Plant Mol. Biol.* **22**, 1067–1085 (1993).
4. Dorer, D.R. & Henikoff, S. Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila. *Cell* **77**, 1–20 (1994).
5. Sabl, J.F. & Henikoff, S. Copy number and orientation determine the susceptibility of a gene to silencing by nearby heterochromatin in Drosophila. *Genetics* **142**, 447–458 (1996).
6. Devis, B.P. & MacDonald, R.J. Limited transcription of rat elastase I transgene repeats in transgenic mice. *Genes Dev.* **2**, 13–22 (1988).
7. Dorer, D.R. Do transgene arrays form heterochromatin in vertebrates? *Transgenic Res.* **5**, 3–10 (1997).
8. Sharpe, J.A. *et al.* Analysis of the human α-globin gene cluster in transgenic mice. *Proc. Natl. Acad. Sci. USA* **90**, 11262–11266 (1993).
9. Higgs, D.R. *et al.* A major positive regulatory region located far upstream of the human α-globin gene locus. *Genes Dev.* **4**, 1588–1601 (1990).
10. Robertson, G. *et al.* Position-dependent variegation of globin transgene expression in mice. *Proc. Natl. Acad. Sci. USA* **92**, 5371–5375 (1995).
11. Hoess, R.H. & Abremski, K. Interaction of the bacteriophage P1 recombinase Cre with the recombining site *loxP*. *Proc. Natl. Acad. Sci. USA* **81**, 1026–1029 (1984).
12. Lakso, M. *et al.* Efficient *in vivo* manipulation of mouse genomic sequences at the zygote stage. *Proc. Natl. Acad. Sci. USA* **89**, 6232–6235 (1992).
13. Palmiter, R.D. & Brinster, R.L. Germ-line transformation of mice. *Annu. Rev. Genet.* **20**, 465–499 (1986).
14. Araki, K., Araki, M., Miyazaki, J.-I. & Vassalli, P. Site-specific recombination of a transgene in fertilized eggs by transient expression of Cre recombinase. *Proc. Natl. Acad. Sci. USA* **92**, 160–164 (1995).
15. Robertson, G., Garrick, D., Wilson, M., Martin, D.I.K. & Whitelaw, E. Age-dependent silencing of globin transgenes in the mouse. *Nucleic Acids Res.* **24**, 1465–1471 (1996).
16. Sutherland, H.G.E., Martin, D.I.K. & Whitelaw, E. A globin enhancer acts by increasing the proportion of erythrocytes expressing a linked transgene. *Mol. Cell. Biol.* **17**, 1607–1614 (1997).
17. Meyer, P., Heidmann, I. & Niedenhof, I. Differences in DNA-methylation are associated with a paramutation phenomenon in transgenic petunia. *Plant J.* **4**, 89–100 (1993).
18. Neuhuber, F., Park, Y.-D., Matzke, A.J.M. & Matzke, M.A. Susceptibility of transgene loci to homology-dependent gene silencing. *Mol. Gen. Genet.* **244**, 230–241 (1994).
19. Ye, F. & Signer, E.R. RIGS (repeat-induced gene silencing) in *Arabidopsis* is transcriptional and alters chromatin configuration. *Proc. Natl. Acad. Sci. USA* **93**, 10881–10886 (1996).
20. de Carvalho, F. *et al.* Suppression of beta-1,3-glucanase transgene expression in homozygous plants. *EMBO J.* **11**, 2595–2602 (1992).
21. Ingelbrecht, I., Van Houdt, H., Van Montagu, M. & Depicker, A. Posttranscriptional silencing of reporter transgenes in tobacco correlates with DNA methylation. *Proc. Natl. Acad. Sci. USA* **91**, 10502–10506 (1994).
22. Bresnick, E.H. & Felsenfeld, G. Dual promoter activation by the human β-globin locus control region. *Proc. Natl. Acad. Sci. USA* **91**, 1314–1317 (1994).
23. Martin, D.I.K. & Whitelaw, E. The vagaries of variegating transgenes. *Bioessays* **18**, 919–923 (1996).
24. Grosveld, F., van Assendelft, G.B., Greaves, D.R. & Kollias, G. Position-independent, high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985 (1987).
25. Ryan, T.M. *et al.* A single erythroid specific DNase I superhypersensitive site activates high levels of human β-globin gene expression in transgenic mice. *Genes Dev.* **3**, 314–323 (1989).
26. Matzke, M.A., Primig, M., Trnovsky, J. & Matzke, A.J.M. Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants. *EMBO J.* **8**, 643–649 (1989).
27. Rhounim, L., Rossignol, J.-L. & Faugeron, G. Epimutation of repeated genes in *Accobulus immersus*. *EMBO J.* **11**, 4451–4457 (1992).
28. Urieli-Shoval, S., Gruenbaum, Y., Sedat, J. & Razin, A. The absence of detectable methylated bases in *Drosophila melenogaster* DNA. *FEBS Lett.* **146**, 148–152 (1982).
29. Garrick, D., Sutherland, H., Robertson, G. & Whitelaw, E. Variegated expression of a globin transgene correlates with chromatin accessibility but not methylation status. *Nucleic Acids Res.* **24**, 4902–4909 (1996).
30. Pondel, M.D., Proudfoot, N.J., Whitelaw, C. & Whitelaw, E. The developmental regulation of the human ζ-globin gene in transgenic mice employing β-galactosidase as a reporter gene. *Nucleic Acids Res.* **20**, 5655–5660 (1992).
31. Whitelaw, E. & Proudfoot, N. α-Thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3′ end processing in the human α2 globin gene. *EMBO J.* **5**, 2915–2922 (1986).

TAB 3

MOLECULAR AND CELLULAR BIOLOGY, Oct. 2000, p. 7419–7426 Vol. 20, No. 20
0270-7306/00/$04.00+0
Copyright © 2000, American Society for Microbiology. All Rights Reserved.

Retroviral Expression in Embryonic Stem Cells and Hematopoietic Stem Cells

SARA R. CHERRY,[1,2] D. BINISZKIEWICZ,[1] L. VAN PARIJS,[3] D. BALTIMORE,[3] AND R. JAENISCH[1,2]*

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142[1]; Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[2]; and California Institute of Technology, Pasadena, California 91125[3]

Received 5 May 2000/Accepted 19 June 2000

Achieving long-term retroviral expression in primary cells has been problematic. De novo DNA methylation of infecting proviruses has been proposed as a major cause of this transcriptional repression. Here we report the development of a mouse stem cell virus (MSCV) long terminal repeat-based retroviral vector that is expressed in both embryonic stem (ES) cells and hematopoietic stem (HS) cells. Infected HS cells and their differentiated descendants maintained long-term and stable retroviral expression after serial adoptive transfers. In addition, retrovirally infected ES cells showed detectable expression level of the green fluorescent protein (GFP). Moreover, GFP expression of integrated proviruses was maintained after in vitro differentiation of infected ES cells. Long-term passage of infected ES cells resulted in methylation-mediated silencing, while short-term expression was methylation independent. Tissues of transgenic animals, which we derived from ES cells carrying the MSCV-based provirus, did not express GFP. However, treatment with the demethylating agent 5-azadeoxycytidine reactivated the silent provirus, demonstrating that DNA methylation is involved in the maintenance of retroviral repression. Our results indicate that retroviral expression in ES cells is repressed by methylation-dependent as well as methylation-independent mechanisms.

Retroviral vectors are appealing vehicles for gene transfer. However, long-term expression mediated by integrated proviruses in primary cells has been difficult to achieve. Retroviral regulatory elements are repressed in numerous cell types, including embryonic stem (ES) cells and hematopoietic stem (HS) cells (1, 3). For example, vectors that are functional in mature hematopoietic cells are often not expressed in blood cells of animals transplanted with the infected stem cells (18, 19, 31). In particular, the lack of significant provirus transcription in ES cells and their differentiated descendants has hampered the use of retroviral vectors in transgenic experiments (5, 12, 32). Interestingly, this block in provirus expression is maintained upon differentiation of infected cells despite the fact that primary infection of cells after differentiation results in efficient expression (6, 7, 26).

Transcriptional repression is thought to be mediated by both *cis*-acting de novo methylation of the integrated proviruses and cell-type-specific *trans*-acting transcriptional repressors (5, 9, 23). The effect of *trans*-acting factors on retroviral expression through binding of specific sequences within the promotors of retroviruses has been examined in many studies (29, 30, 35). In fact, the mouse stem cell virus (MSCV) long terminal repeat (LTR) was generated by the modification of the sequences within the LTR to increase the affinity for positive factors and decrease the affinity for negative regulators (20).

In contrast, the role of methylation in silencing has been less clear. DNA methylation is thought to be a general mechanism used by cells to silence foreign DNA and may be involved in the cell defense against transposable elements (39). DNA methylation has also been associated with the repression of gene expression and the silencing of viral control elements (2, 14, 38). Exogenously introduced retroviruses silenced in vitro and in vivo can be reactivated by treatments that result in genomewide demethylation. In addition, transcriptionally silent endogenous retroviral elements are reactivated upon loss of genomic methylation in *Dnmt1* knockout mice (38). Therefore, DNA methylation is thought to causally repress expression of retroviral promoters in a variety of cell types.

ES cells provide a good model to study the role of DNA methylation in retroviral silencing. First, it was demonstrated that ES cells have high de novo methylation activity, which leads to effective methylation of integrated retroviral vectors, while little or no de novo methylation activity was detected in differentiated cells (21). In addition, ES cells were genetically modified to alter the endogenous level of DNA methylation by the targeted disruption of the maintenance methyltransferase gene *Dnmt1*. ES cells homozygous for this mutation proliferate normally with their genomic DNA highly demethylated, while differentiated cells and mice die due to the loss of genomic methylation (21, 22). Therefore, these modified ES cells are useful to study the effect of DNA methylation on retroviral gene expression. In addition, ES cells can be induced to differentiate in vitro or in vivo, allowing the study of DNA methylation and its effect on long-term expression.

Both Moloney virus-based and MSCV-based retroviral vectors have been used for gene transduction in a variety of cells. The MSCV vector is different from the typical Moloney virus vector in that the mutations in the LTR have allowed expression in a larger host range (8, 20). To this end, we modified MSCV to express the green fluorescent protein (GFP) as a sensitive reporter for gene expression (37). Using this vector, we demonstrated efficient expression in both ES and HS cells. We also demonstrated that silencing of retroviruses involves two mechanisms: (i) *trans*-acting factors that affect the initial expression of Moloney virus-based vectors but not MSCV-based vectors and (ii) long-term DNA methylation-dependent silencing that directly restricts expression of MSCV in ES cells and during embryogenesis. Silencing of the MSCV vector in wild-type ES cells and in in vivo differentiated ES cells was

* Corresponding author. Mailing address: Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142. Phone: (617) 258-5186. Fax: (617) 258-6505. E-mail: jaenisch@wi.mit.edu.

Downloaded from mcb.asm.org by on September 14, 2009

reversed by 5-azadeoxycytidine (5-azadC) treatments that demethylated the retroviral sequences, demonstrating that DNA methylation directly controls the maintenance of retroviral repression.

MATERIALS AND METHODS

Tissue culture. ES cells were cultured as described previously (21). To generate ES cell clones for injection into blastocysts, the ES cells were maintained on irradiated mouse embryonic fibroblasts (MEFs) with 500 U of leukemia inhibitory factor (LIF) per ml (22). For other experiments, the ES cells were cultured without MEFs in 1,000 U of LIF per ml. 293 cells were maintained in Dulbecco's modified Eagle medium supplemented with 10% fetal bovine serum (FBS), penicillin, streptomycin, and glutamine. Abelson virus-transformed B cells were maintained in RPMI 1640 supplemented with 10% defined FBS (HyClone), penicillin, streptomycin, glutamine, and 50 μM β-mercaptoethanol. ES cells with retroviral integrants were in vitro differentiated as follows: the cells were passaged without LIF in the absence of MEFs on bacterial plastic petri dishes for 4 days, trypsinized, and cultivated with or without retinoic acid for 2 weeks (25).

Plasmids. The retroviral vectors MIgGFP, pMXGFP, and MSCViresGFP have been described elsewhere (27, 33, 37). The MSCViresGFP vector was modified by introducing either the Cre recombinase or the human Bcl-2 gene upstream of the internal ribosome entry site (IRES)-GFP cassette as described elsewhere (11, 37). The replication-incompetent helper plasmid pCL-eco was used (24).

Retroviral infections. To generate retroviral supernatants, 293 cells were transiently transfected by calcium phosphate-mediated coprecipitation with 5 μg of the replication-incompetent helper vector pCL-eco and 10 μg of the reporter retroviral vector as stated elsewhere (28). The cells were fed at 24 h postinfection, and the retroviral supernatant was used at 48 h. The cells continued to produce high-titer retroviruses for another 2 days, and that supernatant was used if needed for additional experiments. The supernatant was collected, brought to 4 μg of Polybrene per ml–10 mM HEPES, and filtered (0.45-μm-pore-size filter) for use.

ES cells for infection were washed and trypsinized. They were then plated at 10^6 cells per well of a six-well dish and centrifuged. The ES cell medium was removed, and retroviral supernatant was added at 1 ml/10^6 cells. Next, the plate was centrifuged for 45 min at 2,500 rpm at room temperature. The retroviral supernatants were removed; the cells were resuspended in ES cell medium and plated onto gelatinized dishes. ES cells used to generate mice were plated onto irradiated MEFs.

Bone marrow was infected as follows (36). C57BL/6 mice were obtained from the Jackson Laboratory (Bar Harbor, Maine). Bone marrow cells were harvested from the tibias and femurs of C57BL mice 5 days after they received an intraperitoneal injection of 5 mg of 5-fluorouracil (Sigma) in Dulbecco's phosphate-buffered saline (Gibco/BRL). These cells were then cultured for 4 days at 2×10^6 cells/ml with recombinant mouse interleukin-3 (rmIL-3; 20 ng/ml), rmIL-6 (50 ng/ml), and (50 ng/ml; recombinant mouse stem cell factor; R&D Systems) in Dulbecco's modified Eagle medium containing 10% FBS. After 48 and 72 h, the bone marrow cells were spin infected with the retroviral supernatant generated as described above. Then the retroviral supernatant was removed and replaced with growth medium containing cytokines.

FACS (fluorescence-activated cell sorting) analysis and sorting. Adherent cell lines were trypsinized, washed, and resuspended in complete medium to achieve a single-cell suspension at the time points indicated. Nonadherent cells were used directly for analysis. Organs were disrupted manually and passed through a 70-μm mesh to generate a single cell suspension. The cells were analyzed for viability using scatter properties and the exclusion of propidium iodide. The level of GFP expression was monitored by fluorescence without compensation to detect cells with low levels of GFP expression. The ES cells were sorted into ES cell medium and plated immediately onto either gelatinized plates or MEFs for blastocyst injections. The survival of ES cells after sorting was approximately 50%, as measured by the number of colonies generated divided by the expected number of colonies.

5-AzadC treatments. ES cells were treated with 0.15 μM 5-azadC (Sigma) at days 1 and 3 postplating. The cells were fed, allowed to recover, and then assayed 4 to 8 days later. The red blood cells in whole blood were lysed (5), and the remaining cells were stained with fluorescently labeled anti-H2-b, anti-H2-d, anti-B220, anti-TCRa (Pharmingen) at 1:200 as indicated. At day 0, splenocytes were treated with either anti-CD3 or anti-CD40 (Pharmingen); 0.15 μM 5-azadC was added at day 1, and the anti-CD3-treated cells were assayed at day 4. 5-AzadC was added again to the B-cell cultures with fresh anti-CD40 at day 4, and the cells were assayed at day 6.

Staurosporine-mediated cell death. ES cells were infected with the stated retrovirus and treated with staurosporine at day 4 postinfection for 24 h with the indicated concentration of drug. The percentage of viable, GFP-positive cells was determined by flow cytometry (6). Data are presented as a percentage of GFP-positive cells before treatment. Results from one representative experiment of three performed are shown.

LacZ staining. ES cells were infected with the stated retrovirus and sorted for GFP expression at day 3 postinfection. The ES cells were plated and cultured for an additional 5 days and stained for LacZ expression as described elsewhere (41).

Adoptive transfers. Recipient mice (10) received a total of 1,200 rads of whole-body radiation in two doses (800 and 400 rads) 3 h apart and were then injected with 2×10^6 to 5×10^6 infected bone marrow cells. Irradiated mice were maintained on trimethaprim-sulfamethoxazole in sterile cages for 4 to 6 weeks to prevent opportunistic infections (34). Serial passages were performed by harvesting bone marrow from mice 6 to 8 weeks postreconstitution and transferring 2×10^6 to 5×10^6 cells into irradiated recipients. Mice were analyzed 8 to 12 weeks posttransfer to allow reconstitution of the T-cell compartment. These experiments were repeated multiple times with similar results.

Southern blot analysis. The genomic DNA was isolated as described elsewhere (19). Ten micrograms of DNA was digested with the stated restriction enzyme overnight. The products were resolved on an agarose gel, transferred to a nylon membrane, and detected using a probe that spans the entire GFP coding sequence.

RESULTS

High-efficiency retroviral expression in ES cells. Retroviral vectors based on the MSCV LTR were constructed with a multiple cloning site followed by an IRES driving expression of the gene for GFP as schematically diagrammed in Fig. 1A (MiG) (37). We generated high-titer retroviruses by transient transfection and infected ES cells with an adapted spin infection protocol. Using this protocol, we reproducibly achieved high-efficiency (>50%) infection of ES cells as measured by flow cytometry; uninfected control cells were negative for GFP expression (Fig. 1B). The intracellular concentration of GFP is directly proportional to the fluorescence intensity measured by flow cytometry.

Next, we compared expression of the MSCV-based retrovirus and Moloney virus-based retroviral vectors in ES cells. GFP expression was detectable with the MSCV LTR-containing MiG vector but not with the two Moloney virus-based viruses pMX (27) and Mfg (33) (Fig. 1B). This was not due to inefficient genomic integration of the provirus or to a lower titer. Southern blot analysis of genomic DNA demonstrated that all three proviruses were integrated in the ES cells (Fig. 1C). Also, when parallel B-cell cultures were infected with the retroviral supernatant used to infect ES cells, all of the retroviral vectors were expressed in B cells at comparable efficiencies (Fig. 1B).

GFP expression driven by the MSCV LTR in ES cells was substantially lower than in other differentiated cell lines tested (Fig. 1B and data not shown) (20). To determine whether this low level of expression was sufficient to drive functional expression of other gene products, we cloned the Cre recombinase upstream of the IRES-GFP cassette to generate MSCVCreiresGFP. We tested for Cre activity by infecting ES cells that contain a translational stop sequence flanked by *loxP* sites located between the Rosa 26 promoter and a *lacZ* reporter schematically diagrammed in Fig. 2A (34). If Cre is expressed at functional levels in these ES cells, the protein will catalyze recombination of the *loxP* sites, leading to loss of the stop sequences and the expression of LacZ. Indeed, we found that >99% of GFP-positive ES cells that were infected with the Cre-expressing retrovirus were also LacZ positive (Fig. 2B). Uninfected cells were both GFP negative and LacZ negative (data not shown). This indicates that virus-mediated gene transfer resulted in functional Cre expression.

Because Cre activity is required only transiently for LacZ expression, we tested a second gene product that must be stably expressed throughout the experiment. It has been demonstrated that Bcl-2 expression protects many cell types against staurosporine-mediated apoptosis (10). Therefore, we examined whether Bcl-2 could protect ES cells from apoptosis when expressed from the MSCV LTR. We cloned human Bcl-2 up-

Downloaded from mcb.asm.org by on September 14, 2009

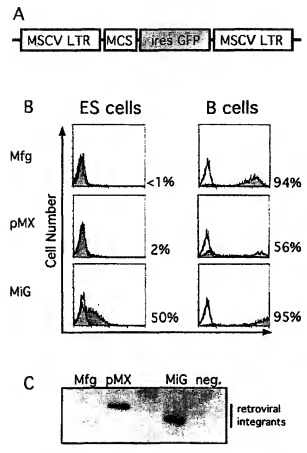


FIG. 1. Efficient retroviral infection of ES cells. (A) Schematic diagram of MiG vector containing the MSCV LTR followed by a multiple cloning site (MCS) and an IRES-GFP cassette. (B) MSCV-based (MiG) but not Moloney virus-based (Mfg and pMX) retroviruses express in ES cells. B cells or ES cells were infected by the indicated retroviruses and assayed by flow cytometry 2 days postinfection. Uninfected cells (unshaded) and infected cells (shaded area) were electronically gated for live cells and subsequently analyzed for GFP fluorescence and for cell number. Percentages of GFP-positive cells are indicated. (C) Comparable levels of integration of different retroviruses into ES cells, determined by Southern blot analysis of genomic DNA purified from infected and uninfected ES cells 2 days postinfection, digested with *Kpn*I, a restriction site present within the LTRs, and probed with the GFP coding sequence.

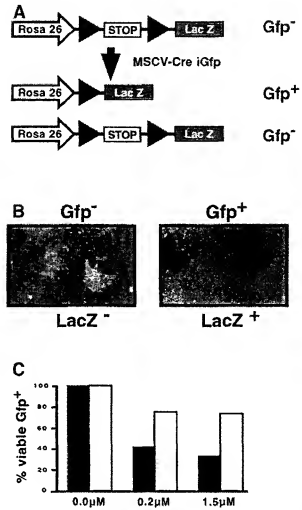


FIG. 2. Expression from the MSCV LTR is sufficient to drive functional gene expression. (A) Schematic diagram of the Rosa 26 locus in Cre reporter ES cells. Before Cre-mediated recombination, LacZ expression is prevented by the presence of a stop fragment. Retroviral infection with a Cre-expressing retrovirus with a GFP reporter results in two populations of cells. Cells that are GFP$^+$ become LacZ$^+$ due to efficient Cre-mediated recombination of the stop fragment. In contrast, cells that are GFP$^-$ were not infected and thus remained LacZ$^-$. (B) ES cells were infected with the MSCVCreiresGFP retrovirus and sorted for either Gfp$^-$ or Gfp$^+$ as indicated. The cells were subsequently cultured and stained for LacZ expression. Gfp$^-$ cells are white (and therefore LacZ$^-$) while Gfp$^+$ cells are blue (and therefore LacZ$^+$). More than 99% of the Gfp$^+$ cells were LacZ$^+$ in multiple experiments. (C) ES cells were infected with either MSCViresGFP (■) or MSCVBcl-2iresGFP (□) and treated with the indicated amounts of staurosporine. The percentage of viable, Gfp$^+$ (infected) cells was determined by flow cytometry. The results are shown as a percentage of Gfp$^+$ cells before treatment. The results are from one representative experiment of three performed.

stream of the IRES-GFP cassette to generate MSCV Bcl-2iresGFP. Wild-type ES cells were infected with either the Bcl-2-expressing retrovirus or the control virus lacking Bcl-2. Increasing concentrations of staurosporine were added to the cultures, and flow cytometry was used to assay for both viability and GFP expression. GFP-positive cells infected with the Bcl-2-containing virus were significantly protected from staurosporine-mediated cell death compared to the GFP-negative cells or GFP-positive cells infected with the control retrovirus (Fig. 2C). Therefore, the level of expression from the MSCV LTR is sufficient for stable functional gene expression in ES cells.

Short-term transcriptional silencing in ES cells is methylation independent. It long has been hypothesized that retroviruses are transcriptionally silenced in embryonic cells by DNA methylation (12, 14, 21). Therefore, it was possible that DNA methylation of the MSCV LTR was responsible for the decreased level of expression in ES cells compared to other cell types (Fig. 1B). In addition, we sought to test whether DNA methylation of the Moloney virus-based vectors in the wild-type ES cells was the mechanism by which the Moloney virus-based LTRs were silenced (9, 13). To this end, we infected ES cells deficient for the maintenance DNA methyltransferase gene, *Dnmt1*, the loss of which results in genomewide hypomethylation (21, 22). *Dnmt1*$^{-/-}$ ES cells are demethylated, and proviral sequences remain unmethylated. The Moloney virus-based retroviruses such as pMX remained silent even when

Downloaded from mcb.asm.org by on September 14, 2009

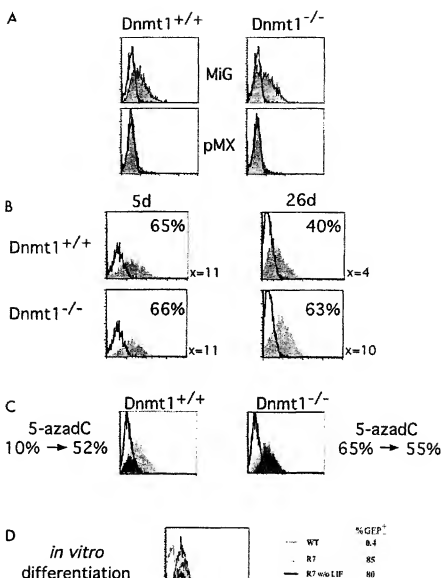


FIG. 3. Long-term expression of retroviruses is repressed by methylation. (A) MSCV-based (MiG) but not Moloney virus-based (pMX) retroviruses express in ES cells independent of the methylation status of the cells. *Dnmt1*$^{+/+}$ or *Dnmt1*$^{-/-}$ ES cells were not infected (unshaded) or infected by the indicated retroviruses (shaded) and assayed by flow cytometry 2 days postinfection as for Fig. 1B. (B) Long-term expression of GFP in ES cells is decreased in *Dnmt1*$^{+/+}$ cells but not *Dnmt1*$^{-/-}$ cells. The ES cells were infected with MiG, passaged for 5 or 26 days postinfection, and assayed by flow cytometry as above. The mean fluorescent intensity for the population and the percentage of GFP-positive cells are indicated. (C) Treatment with 5-azadC rescues the expression of retroviruses silenced by long-term passage. *Dnmt1*$^{+/+}$ or *Dnmt1*$^{-/-}$ ES cells were infected by MiG and passaged for >40 days. The cells were divided, and half were treated with 5-azadC. Then uninfected ES cells (unshaded), MiG-infected untreated ES cells (dark shading), and MiG-infected 5-azadC-treated ES cells (light shading) were assayed by FACS analysis. Numbers below the FACS plots are percentages of GFP-positive cells before and after 5-azadC treatment. (D) In vitro differentiation of ES cells does not affect retroviral expression. A clonal ES cell line (R7) infected with MiG or an uninfected wild-type (WT) ES cell control was in vitro differentiated by passage without feeders and LIF, with or without retinoic acid (RA) as indicated. The cells were assayed by flow cytometry, and the percentage of GFP-positive cells is indicated.

introduced into *Dnmt1*$^{-/-}$ ES cells, whereas MSCV expressed similar levels of GFP in both *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ ES cells (Fig. 3A). Therefore, the initial block in transcription directed by Moloney virus LTRs in ES cells is independent of DNA methylation and is presumably due to the binding of *trans*-acting factors. In addition, the mean fluorescence intensities of GFP were comparable between the *Dnmt1*$^{+/+}$ and *Dnmt1*$^{-/-}$ ES cells, indicating that the basal level of expression of the MSCV LTR is independent of DNA methylation.

DNA methylation constrains long-term retroviral expression. MiG-infected GFP-expressing ES cells were continually passaged to test the effect of DNA methylation on long-term expression.

Downloaded from mcb.asm.org by on September 14, 2009

Though GFP expression was high in both *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ ES cells at 5 days postinfection, a substantial fraction of the infected wild-type ES were GFP negative at 26 days postinfection. This was apparent by both a loss in the percentage of GFP-positive cells as well as a decrease in the mean fluorescence intensity of the bulk population of wild-type ES cells and was observed in both bulk cultures and individual cloned lines containing single integrants (Fig. 3B and data not shown). The fraction of GFP-positive cells continues to decrease with additional passages, as shown in Fig. 3C. These results suggest that long-term expression was suppressed by DNA methylation. To directly test whether retroviral repression was due to de novo methylation of the newly integrated retroviruses, we treated the long-term cultures with 5-azadC, a drug that leads to hypomethylation of genomic DNA (16). If DNA methylation was preventing expression of the MSCV LTR, treatment with the drug should activate retroviral expression. Indeed, we found that 5-azadC treatment of ES cells that had lost expression of GFP through long-term passage reactivated the provirus (Fig. 3C). In contrast, *Dnmt1*$^{-/-}$ ES cells infected with the retrovirus did not lose expression of GFP; thus, treatment with 5-azadC did not significantly affect retroviral expression (Fig. 3C). We also analyzed clonal lines containing single proviral integrants in which GFP expression was progressively silenced and found that treatment with 5-azadC resulted in the reactivation of gene expression in all cases (data not shown). This demonstrates that DNA methylation controls long-term but not short-term expression of retroviruses in ES cells.

Expression is maintained after in vitro differentiation. Previously, in vitro differentiation of ES cells had been demonstrated to silence expression of retroviral sequences (12, 20). Thus, we tested whether GFP expression from the MiG retrovirus in ES cells was affected by in vitro differentiation. We cultured MiG-infected wild-type ES cells in the absence of embryonic feeder cells and LIF in suspension to generate embryoid bodies. Disaggregated embryoid bodies were replated either with or without retinoic acid. We found no change in GFP expression in MiG-infected bulk cultures or individual subclones containing one to several integrants upon in vitro differentiation with either method, as shown for one clonal line containing multiple integrants in Fig. 3D. GFP expression was unchanged in all in vitro-differentiated ES cell lines, regardless of whether the subclones contained only a single or multiple integrants. This indicates that the MSCV-based MiG retrovirus is not silenced by in vitro differentiation.

Generation of mice from GFP-expressing MiG-infected ES cells. We next determined whether expression of the MSCV-based MiG vector was affected by in vivo differentiation of the infected ES cells. Cells from the chimeric animals were derived by injection of MiG-infected wild-type ES cells (derived from 129-Sv/Jae mice) into BALB/c blastocysts. MiG-infected *Dnmt1*$^{-/-}$ ES cells cannot be used for injection into blastocysts, because *Dnmt1*$^{-/-}$ ES cells die upon differention and therefore do not contribute significantly to adult mice (22). MiG-infected wild-type ES cells were sorted for GFP expression by flow cytometry prior to injection, and two GFP-expressing clones, R2 and R11, were isolated (Fig. 4B). Southern blot analysis demonstrated that R2 contained two integrants that comigrate on an agarose gel, and R11 contained three proviral integrants (Fig. 4A). High-contribution chimeras (>80% by coat color) were generated from the R2 and R11 ES cells, which transmitted the proviruses to their offspring (data not shown).

To test whether the chimeras expressed the integrated retroviruses, we isolated peripheral blood mononuclear cells

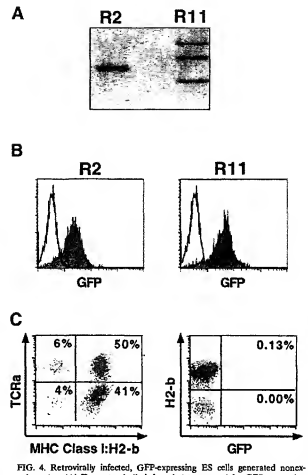


FIG. 4. Retrovirally infected, GFP-expressing ES cells generated nonexpressing mice. (A) Two retrovirally infected clones sorted for GFP expression were analyzed for proviral integrants by Southern blot analysis. R2 contained two integrants, while R11 contained three. Uninfected cells are negative. (B) The clones were passaged after sorting for GFP-expressing cells by flow cytometry and reanalyzed for GFP expression. Both R2 and R11 express GFP (shaded) compared to uninfected controls (unshaded). (C) PBMCs from the R2 chimera (more than 50% contribution by coat color) were analyzed by flow cytometry. ES cell contribution to the chimera was determined by phycoerythrin-H2-b staining and cyc-TCRa staining and demonstrated contribution to the T-cell compartment. The percentage of cells in each quadrant is listed. The cells were also monitored for GFP expression. The percentage of GFP$^+$ cells that are either major histocompatibility complex (MHC) class I H2-b$^+$ (ES cell derived) or H2-b$^-$ (blastocyst derived) is listed in the quadrant.

(PBMCs) from both the R2- and R11-derived chimeras. To distinguish whether the PBMCs were derived from the ES cell donor or the host blastocyst, we stained the cells with antibodies that recognized specific major histocompatibility complex class I haplotypes (Pharmingen). The donor ES cells (129 derived) are H2-b, and the blastocysts (BALB/C derived) are H2-d (Fig. 4C and data not shown). In addition, we stained the PBMCs with a pan-T-cell (TCRa) (Fig. 4C) or pan-B-cell (B220) antibody (data not shown) to determine the ES cell contribution to these lineages. Using this strategy, we found that approximately 90% of the PBMCs from either the R2 or R11 chimera were ES cell derived as measured by H2-b staining (Fig. 4C, data not shown). However, the majority of the cells did not express GFP in either chimera (Fig. 4C and data not shown). On the order of 0.1% of the PBMCs that were ES

Downloaded from mcb.asm.org by on September 14, 2009

cell derived were GFP positive, compared to less than 0.01% that were blastocyst derived (Fig. 4C). Similar results were also obtained with cells from the R11 chimera (data not shown). The results indicate that the MSCV LTR is repressed during in vivo differentiation to lymphocytes. Nevertheless, a small number of cells escaped silencing and expressed GFP. This transcriptional repression of the MiG provirus in the chimeras is in contrast to the GFP expression both in the donor ES and after in vitro differentiation (Fig. 4B, data not shown).

To determine if other somatic cells expressed the retroviral integrants, we analyzed the progeny of the chimeras. We isolated spleen, thymus, kidney, and liver cells from an animal carrying the two proviral integrants present in the R2 chimera and a littermate control containing no retroviral integrants. We analyzed these cells for GFP expression by FACS analysis and found no detectable expression of GFP in the splenocytes, thymocytes, renal cells, or hepatoces (Fig. 5A and data not shown).

In vitro reactivation of retroviral expression. One possible explanation for transcriptional repression during in vivo differentiation was de novo methylation of the integrated retroviral LTR during embryonic development. To test this hypothesis, we cultured splenocytes from a mouse containing the R2 proviruses and from a littermate control, by treating the cells with either anti-CD3 or anti-CD40 to activate and induce proliferation of the T cells or B cells, respectively (4). We then assayed for GFP expression by flow cytometry and found that proliferation of the splenocytes did not activate expression of the retrovirus (data not shown). Next, we added 5-azadC to the splenocyte cultures to induce demethylation of the retroviruses. Indeed, treatment with 5-azadC activated expression in approximately 2% of the T cells (anti-CD3) (Fig. 5B) and 2% of the B cells (anti-CD40) (data not shown). In addition, when in vivo-differentiated cells, which had been isolated from the kidney of a transgenic mouse and transformed with simian virus 40 large T antigen (15), were treated with 5-azadC, activation of the silent provirus was observed in a similar fraction of the cells (data not shown). The extent of reactivation of expression of the provirus in in vivo-differentiated cells by 5-azadC was lower than in ES cells, where the reactivation of the provirus with 5-azadC was almost complete.

We next determined whether demethylation of the retrovirus in vivo would activate expression of the integrated retroviruses (13). Newborn mice were subcutaneously injected with 5-azadC at postnatal day 5 and subsequently analyzed at postnatal day 14 for GFP expression by flow cytometry. We found that 5-azadC-injected animals but not the uninjected controls had activated GFP expression of the proviruses in the spleen, thymus, and kidney (Fig. 5C and data not shown). When we injected higher concentration of 5-azadC in an effort to further demethylate the newborn mice, all injected animals died. This result demonstrated that repression by DNA methylation is, at least in part, responsible for silencing expression of the retroviral LTR in vivo.

Retroviral expression in HS cells after serial adoptive transfers. Bone marrow contains the HS cells that can stably repopulate the hematopoietic system after transfer to lethally irradiated mice. To determine whether HS cells can be effectively transduced and express the MiG retrovirus, we used infected bone marrow cells to reconstitute lethally irradiated mice (Fig. 6). We found that between 30 and 80% of the splenocytes from these primary recipients expressed the retrovirus, as measured by FACS analysis for GFP expression and shown for one representative experiment (Fig. 6A). The MiG virus was expressed in the B-cell, T-cell, and granulocyte compartments, as measured by a pan-B cell (B220), pan-T-cell (Thy-1), and pan-granulocyte (Gr-1) marker electronically gated on GFP-positive cells (Fig. 6B and data not shown). Because a large fraction of the splenocytes in the primary recipients are derived from relatively differentiated, lineage-committed progenitors, serial adoptive transfers are required to test for retroviral expression in the true HS cells (17). Therefore, we used bone marrow from these primary recipients to serially reconstitute lethally irradiated mice. This protocol requires substantial expansion from the stem cells and tests for long-term expression of the retrovirus. We observed no change in the percentage GFP-positive HS cells, and the level of GFP expression from the adoptive transfers into multiple recipients was stable over three additional passages (4° recipient). In addition, the infected cells gave rise to both B- and T-cell lineages at the expected ratios (Fig. 6B), demonstrating not only that the MiG retrovirus transduced the long-term repopulating HS cells but also that the MiG-mediated GFP expression was stable during in vivo hematopoietic differentiation. However, our results do not exclude the possibility that in addition to the transcriptionally active proviruses present within these cells, there are also copies of the virus that were transcriptionally silenced.

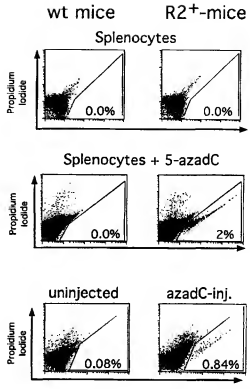


FIG. 5. Silenced retroviruses can be reactivated with 5-azadC. (A) Splenocytes from an R2$^+$ or R2$^-$ littermate do not express GFP by flow cytometry. Cells were stained with propidium iodide to exclude dead cells, and the percentage of GFP$^+$ cells is indicated. (B) The splenocytes from panel B were induced to proliferate with anti-CD3 and treated with 5-azadC. Cells were stained with propidium iodide to exclude dead cells and analyzed by flow cytometry. The percentage of GFP$^+$ cells is indicated. (C) Flow cytometric analysis of the splenocytes of littermates that were either uninjected or injected with 5-azadC at passage 5 and analyzed at passage 14 for GFP expression. The percentage of GFP$^+$ splenocytes is indicated.

Downloaded from mcb.asm.org by on September 14, 2009

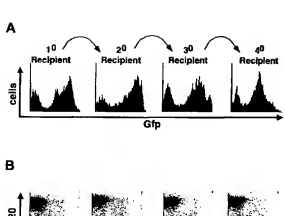


FIG. 6. Serial adoptive transfers maintain expression of the MSCV-based retrovirus. (A) Bone marrow was infected with MiG and used to reconstitute multiple lethally irradiated mice to generate the 1° recipient. The spleen of the 1° recipient was analyzed for GFP expression by flow cytometry. The bone marrow of the 1° recipient was used to reconstitute lethally irradiated 2° recipients. The spleen of a 2° recipient was analyzed for GFP expression, and the bone marrow was used to reconstitute lethally irradiated 3° recipients. The spleen of a 3° recipient was analyzed for GFP expression, and the bone marrow was used to reconstitute lethally irradiated 4° recipients. A representative analysis is shown. (B) splenocytes from panel A, stained with pan-B-cell (B220) and pan-T-cell (Thy-1) antibodies and electronically gated for GFP$^+$ cells, are shown below the GFP histogram they are derived from. The FACS diagrams are shown for these serially reconstituted spleens, demonstrating that the transferred cells contribute to both B- and T-cell lineages in the appropriate ratios.

DISCUSSION

We have investigated the role of DNA methylation in retroviral silencing. Retrovirus-based studies of stem cells have been hampered by the lack of expression. We have overcome the transcriptional repression in ES cells by using an MSCV-based vector in combination with a sensitive GFP reporter gene (MiG vector). The analysis of expression of the MiG vector and other Moloney virus-based vectors in *Dnmt1*$^{-/-}$ and *Dnmt1*$^{+/+}$ ES cells has allowed us to determine whether DNA methylation directly controls retroviral gene expression in these cells. We found that both methylation-dependent and methylation-independent mechanisms exist to control retroviral gene expression.

Historically, retroviral expression of Moloney virus-based vectors in ES cells has been negligible. In contrast, the MSCV LTR not only transduces GFP expression in ES cells but also expresses other exogenous gene products such as the Cre recombinase and the antiapoptotic factor Bcl-2 at detectable level in ES cells. Therefore, the MSCV LTR can be used to express various transgenes in ES cells and their differentiated descendant cells.

It had been proposed that DNA methylation has evolved as a cellular mechanism to silence retroviral elements, preventing the spread of transposable elements through the genome (39). Indeed, de novo methylation of integrated proviral sequences has been observed in wild-type ES cells, which was correlated with the transcriptional silencing of the retrovirus (14). Our findings are the first demonstration that inhibition of the *Dnmt1* methyltransferase gene prevents silencing of the retroviruses in ES cells. This result provides direct evidence that DNA methylation is causally involved in long-term retroviral repression. Consistent with this conclusion is the demonstration that the transcriptionally silenced proviruses present in long-term *Dnmt1*$^{+/+}$ ES cell cultures can be reversed by treatments with 5-azadC.

In contrast, methylation-independent mechanisms determine initial retroviral expression in ES cells. Wild-type or *Dnmt1*$^{-/-}$ ES cells infected with Moloney virus-based vectors were transcriptionally silent, and therefore this silencing was independent of the DNA methylation status of the cells. Moreover, the basal level of expression from the MSCV-based vector was unaffected by the methylation status of the cells. This formally demonstrates that DNA methylation-independent mechanisms control initial retroviral gene expression in ES cells. Because the basal level of expression of the MSCV LTR in ES cells is lower than in differentiated cell types and not affected by the methylation status of the ES cells, *trans*-acting factors must regulate the initial level of expression.

Previous studies found that retroviruses, including the MSCV LTR, are silenced by the in vitro differentiation process (20). In contrast, we found for the first time that expression of this MSCV-based retrovirus in ES cells was maintained after in vitro differentiation with and without retinoic acid. We were also able to show long-term, stable GFP expression from the MiG vector in HS cells and their differentiated derivatives. MiG-mediated GFP expression from HS cells was stable through serial adoptive transfers, and the HS cells gave rise to GFP-expressing B- and T-cell lineages. Therefore, this MSCV-based retroviral transduction system should allow for a molecular analysis of stem cell biology and differentiation programs by forced expression of exogenous gene products.

It has been postulated that methylation-dependent mechanisms repress retroviral gene expression upon in vivo differentiation (13, 20). To test this, we injected GFP-expressing undifferentiated ES cells into recipient blastocysts and generated chimeric mice. Differentiated tissues derived from these in vivo-differentiated ES cells, such as PBMCs, lacked significant GFP expression. Treatment of ES cell-derived differentiated cells with 5-azadC in vitro or in vivo led to partial reactivation of expression of the silenced retroviruses in lymphoid and nonlymphoid tissues. We conclude from these results that the maintenance of retroviral silencing in vivo involves DNA methylation. However, only a small fraction of the 5-azadC-treated cells reactivated GFP expression, unlike the long-term ES cell cultures, in which every cell reactivated GFP expression. This suggests that methylation-independent mechanisms exist to suppress retroviral expression. Alternatively, 5-azadC treatment of differentiated cells, in contrast to ES cells, may not lead to a level of genomic demethylation sufficient for complete retroviral reactivation. The transgenic animals carrying the silenced MiG proviruses will be a valuable indicator for in vivo activation of GFP expression under different conditions.

ACKNOWLEDGMENTS

The first two authors, S. R. Cherry and D. Biniszkiewicz, contributed equally to this work.

We thank members of the Baltimore and Jaenisch labs for advice and discussions. We are grateful to Brian Bates for discussions and comments on the manuscript, George Daley for discussions, Jessie Dausman for blastocyst injections, Ruth Curry for mouse work, and Glen Paradis for FACS sorting.

D. Biniszkiewicz was supported by the Deutsche Akademische Austauschdienst. D. Baltimore and R. Jaenisch (NIH/NCI 5-R35-CA44339) are supported by NIH grants. This work was supported in part by the ERC program of the National Science Foundation under award EEC-9843342.

REFERENCES

1. **Asche, W., G. Colletta, G. Warnecke, F. Nobis, S. Pennie, R. M. King, and W. Ostertag.** 1984. Lack of retrovirus gene expression in somatic cell hybrids of

Downloaded from mcb.asm.org by on September 14, 2009

Friend cells and teratocarcinoma cells with a teratocarcinoma phenotype. Mol. Cell. Biol. 4:923–930.
2. Cedar, H. 1988. DNA methylation and gene activity. Cell 53:3–4.
3. Challita, P. M., and D. B. Kohn. 1994. Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation in vivo. Proc. Natl. Acad. Sci. USA 91:2567–2571.
4. Colligan, J. E., et al. (ed.). Current protocols in immunology. J. Wiley & Sons, New York, N.Y.
5. Gautsch, J. W. 1980. Embryonal carcinoma stem cells lack a function required for virus replication. Nature 285:110–112.
6. Gautsch, J. W., and M. C. Wilson. 1983. Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature 301:32–37.
7. Greiser-Wilke, I., W. Ostertag, P. Goldfarb, A. Lang, M. Furusawa, and J. F. Conscience. 1981. Inducibility of spleen focus-forming virus by BrdUrd is controlled by the differentiated state of the cell. Proc. Natl. Acad. Sci. USA 78:2995–2999.
8. Grez, M., E. Akgun, F. Hilberg, and W. Ostertag. 1990. Embryonic stem cell virus, a recombinant murine retrovirus with expression in embryonic stem cells. Proc. Natl. Acad. Sci. USA 87:9202–9206.
9. Hoeben, R. C., A. A. Migchielsen, R. C. van der Jagt, H. van Ormondt, and A. J. van der Eb. 1991. Inactivation of the Moloney murine leukemia virus long terminal repeat in murine fibroblast cell lines is associated with methylation and dependent on its chromosomal position. J. Virol. 65:904–912.
10. Huang, D. C., S. Cory, and A. Strasser. 1997. Bcl-2, Bcl-XL and adenovirus protein E1B19kD are functionally equivalent in their ability to inhibit cell death. Oncogene 14:405–414.
11. Jacob, J., and D. Baltimore. 1999. Modelling T-cell memory by genetic marking of memory T cells in vivo. Nature 399:593–597.
12. Jaenisch, R., H. Fan, and B. Croker. 1975. Infection of preimplantation mouse embryos and of newborn mice with leukemia virus: tissue distribution of viral DNA and RNA and leukemogenesis in the adult animal. Proc. Natl. Acad. Sci. USA 72:4008–4012.
13. Jaenisch, R., A. Schnieke, and K. Harbers. 1985. Treatment of mice with 5-azacytidine efficiently activates silent retroviral genomes in different tissues. Proc. Natl. Acad. Sci. USA 82:1451–1455.
14. Jahner, D., and R. Jaenisch. 1985. Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. Nature 315:594–597.
15. Jat, P. S., C. L. Cepko, R. C. Mulligan, and P. A. Sharp. 1986. Recombinant retroviruses encoding simian virus 40 large T antigen and polyomavirus large and middle T antigens. Mol. Cell. Biol. 6:1204–1217.
16. Jones, P. A. 1985. Altering gene expression with 5-azacytidine. Cell 40:485–486.
17. Jones, R. J., J. E. Wagner, P. Celano, M. S. Zicha, and S. J. Sharkis. 1990. Separation of pluripotent haematopoietic stem cells from spleen colony-forming cells. Nature 347:188–189.
18. Kohn, D. B. 1995. The current status of gene therapy using hematopoietic stem cells. Curr. Opin. Pediatr. 7:56–63.
19. Krall, W., and D. B. Kohn. 1996. Expression levels by retroviral vectors based upon the N2 and the MFG backbones. Gene Ther. 3:365.
20. Laker, C., J. Meyer, A. Schopen, J. Friel, C. Heberlein, W. Ostertag, and C. Stocking. 1998. Host *cis*-mediated extinction of a retrovirus permissive for expression in embryonal stem cells during differentiation. J. Virol. 72:339–348.
21. Lei, H., S. P. Oh, M. Okano, R. Juttermann, K. A. Goss, R. Jaenisch, and E. Li. 1996. De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 122:3195–3205.
22. Li, E., T. H. Bestor, and R. Jaenisch. 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69:915–926.
23. Loh, T. P., L. L. Sievert, and R. W. Scott. 1990. Evidence for a stem cell-specific repressor of Moloney murine leukemia virus expression in embryonal carcinoma cells. Mol. Cell. Biol. 10:4045–4057.
24. Naviaux, R. K., E. Costanzi, M. Haas, and I. M. Verma. 1996. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. J. Virol. 70:5701–5705.
25. Nichols, J., E. P. Evans, and A. G. Smith. 1990. Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. Development 110:1341–1348.
26. Niwa, O., Y. Yokota, H. Ishida, and T. Sugahara. 1983. Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. Cell 32:1105–1113.
27. Onishi, M., S. Kinoshita, Y. Morikawa, A. Shibuya, J. Phillips, L. L. Lanier, D. M. Gorman, G. P. Nolan, A. Miyajima, and T. Kitamura. 1996. Applications of retrovirus-mediated expression cloning. Exp. Hematol. 24:324–329.
28. Pear, W. S., G. P. Nolan, M. L. Scott, and D. Baltimore. 1993. Production of high-titer helper-free retroviruses by transient transfection. Proc. Natl. Acad. Sci. USA 90:8392–8396.
29. Petersen, R., G. Kempler, and E. Barklis. 1991. A stem cell-specific silencer in the primer-binding site of a retrovirus. Mol. Cell. Biol. 11:1214–1221.
30. Prince, V. E., and P. W. Rigby. 1991. Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. J. Virol. 65:1803–1811.
31. Robbins, P. B., D. C. Sinclton, X. J. Yu, S. Haiene, E. H. Leonard, and D. B. Kohn. 1998. Consistent, persistent expression from modified retroviral vectors in murine hematopoietic stem cells. Proc. Natl. Acad. Sci. USA 95: 10182–10187.
32. Robertson, E., A. Bradley, M. Kuehn, and M. Evans. 1986. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323:445–448.
33. Silver, D. P., E. Spanopoulou, R. C. Mulligan, and D. Baltimore. 1993. Dispensable sequence motifs in the RAG-1 and RAG-2 genes for plasmid V(D)J recombination. Proc. Natl. Acad. Sci. USA 90:6100–6104.
34. Soriano, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat. Genet. 21:70–71.
35. Tsukiyama, T., H. Ueda, S. Hirose, and O. Niwa. 1992. Embryonal long terminal repeat-binding protein is a murine homolog of PTZ-F1, a member of the steroid receptor superfamily. Mol. Cell. Biol. 12:1286–1291.
36. Van Parijs, L., Y. Refaeli, A. K. Abbas, and D. Baltimore. 1999. Autoimmunity as a consequence of retrovirus-mediated expression of C-FLIP in lymphocytes. Immunity 11:763–770.
37. Van Parijs, L., Y. Refaeli, J. D. Lord, B. H. Nelson, A. K. Abbas, and D. Baltimore. 1999. Uncoupling IL-2 signals that regulate T cell proliferation, survival, and Fas-mediated activation-induced cell death. Immunity 11:281–288.
38. Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. 20:116–117.
39. Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13:335–340.

Downloaded from mcb.asm.org by on September 14, 2009

TAB 4

Gene, 91 (1990) 179–184
Elsevier

This material may be protected by Copyright law (Title 17 U.S. Code)

GENE 03596

The methylation-free status of a housekeeping transgene is lost at high copy number

(Recombinant DNA; hypomethylation; *HTF* island; hydroxy-methylglutaryl CoA reductase; cholesterol)

Majid Mehtali[a,b], Marianne LeMeur[a] and Richard Lathe[a,c]

[a] *LGME-CNRS and U184-INSERM, 11 rue Humann, 67085 Strasbourg Cedex (France) Tel. 88-37 12 55;* [b] *Transgene SA, 11 rue de Molsheim, 67000 Strasbourg (France) Tel. 88 22 24 90;* [c] *AFRC Centre for Animal Genome Research, University of Edinburgh, King's Buildings, West Mains Road, Edinburgh EH9 3JQ (U.K.) Tel. 31 668 4541*

Received by J.-P. Lecocq: 21 August 1989
Revised: 21 January 1990
Accepted: 22 February 1990

SUMMARY

Transgenic mouse lines were established bearing tandem arrays of a fusion construct comprising the promoter region of a housekeeping gene, *HMGCR*, encoding 3-hydroxy 3-methylglutaryl CoA reductase, linked to a bacterial *cat* reporter gene encoding chloramphenicol acetyltransferase (CAT). CAT activity was observed in all transgenic mouse tissues examined. The methylation state of the fusion transgene was determined. In non-transgenic mice the endogenous *HMGCR* promoter is devoid of methylation while flanking regions are extensively modified. In *HMGCR-cat* transgenic mice the fusion gene promoter was found to be similarly hypomethylated. However, the extent of hypomethylation varied with copy number: methylation-free status was progressively lost with increasing transgene copy number. Further transgenic mouse lines were constructed carrying a truncated *HMGCR* regulatory region linked to *cat*. Transgene expression and hypomethylation were observed in testis but not in any other tissue, and testis-specific methylation-free status was also lost at high copy number. Loss of hypomethylation at high copy number may indicate that saturable DNA-binding factors normally protect the *HMGCR* promoter from methylation.

INTRODUCTION

In contrast to the largely methylation-free genomes of bacteria and invertebrates, vertebrate genomes have extensive 5-methylation of C residues within CpG doublets (Grippo et al., 1968), and under-representation of CpG has been attributed to spontaneous deamination of 5-methyl-C (Bird, 1980). The relative insensitivity of vertebrate DNA to cleavage with enzymes whose recognition sequence contains a CpG doublet, for instance *Hpa*II (CCGG), is a consequence both of the rarity of CpG doublets and of cleavage inhibition by methylation of the recognition sequence. However, a small fraction of the genome is efficiently cleaved by *Hpa*II (Cooper et al., 1983) and such regions, '*Hpa*II-tiny-fragment' (HTF) islands, rich in unmethylated CpG doublets, appear to be associated with actively transcribed genes (Lavia et al., 1987) and particularly with the promoter regions of housekeeping genes (Bird, 1986; Gardiner-Garden and Frommer, 1987). Promoter regions of ubiquitously-expressed genes are unusual in that they are G/C-rich and lack consensus transcription initiation signals such as the 'TATA' or 'CAAT' boxes, and ubiquitous expression has been attributed to hypomethylation of control regions. (reviewed by Bird, 1986; 1987; Dynan, 1986; Gardiner-Garden and Frommer, 1987; Cedar, 1988). Methylation can correlate with diminished expression in vivo and in vitro (reviewed by Cedar, 1988);

Correspondence to: Dr. R. Lathe, CAGR, King's Buildings, West Mains Road, Edinburgh EH9 3JQ (U.K.) Tel. 31 668 4541; Fax 31 667 0164.

Abbreviations: bp, base pair(s); CAT, Cm acetyltransferase; *cat*, gene encoding CAT; Cm, chloramphenicol; HMGCR, 3-hydroxy-3-methylglutaryl CoA reductase; *HMGCR*, gene encoding HMGCR; *HTF*, *Hpa*II tiny fragment; kb, kilobase(s) or 1000 bp; nt, nucleotide(s).

0378-1119/90/$03.50 © 1990 Elsevier Science Publishers B.V. (Biomedical Division)

however, the causal link between hypomethylation and gene expression is unclear.

HMGCR (3-hydroxy-3-methylglutaryl CoA reductase) is a membrane-bound glycoprotein that catalyses a key step in the synthesis of cholesterol, an essential component of the cell membrane (Brown and Goldstein, 1980; Luskey, 1986). The *HMGCR* genes of hamster and human are characterized by a noncoding first exon and a long (ca. 3.5 kb) intervening sequence prior to the translation start site in the second exon (Reynolds et al., 1985; Luskey, 1987). We recently isolated the homologous mouse *HMGCR* gene and the promoter region, as for the hamster and human genes, bears the hallmarks of a *HTF* island in high G/C content and a CpG/GpC ratio close to unity (M.M. and R.L., unpublished).

We have used the transgenic mouse system to explore *HTF* island hypomethylation. Pronuclear microinjection of DNA into fertilized mouse eggs is now an established technique for generating mice bearing new gene combinations. In the majority of cases the injected DNA integrates at a single site as a multiple tandem repeat, and transgene copy number varies considerably between different transgenic lines (reviewed by Palmiter and Brinster, 1986). To explore possible variation in transgene functional status with copy number, two different fusion genes between the *HMGCR* promoter region and a *cat* reporter gene were constructed and introduced into the mouse germline. Employing these transgenic animals we endeavoured to study the relationship between transgene copy number, expression, and hypomethylation of the transgene promoter.

RESULTS AND DISCUSSION

(a) Transgenic animals bearing the *HMG1-cat* fusion construct

Construct *HMG1-cat* comprises a 5.5-kb *Bam*HI fragment of the mouse *HMGCR* promoter region, containing 1.35 kb of upstream sequence, the first (noncoding) exon and the first intron, linked to *cat* (Fig. 1A). The fusion construct was injected into fertilized mouse eggs and lines of transgenic mice were obtained carrying between 10 and 260 tandem copies of the transgene. These animals all express the fusion transgene in all tissues examined irrespective of transgenic line (Table I; also not presented): no proportionality was observed between expression level and copy number (M.M. and R.L., unpublished) as recorded (Palmiter and Brinster, 1986) in other transgenic systems.

(b) Methylation status of the *HMG1-cat* transgene

To assess the in vivo methylation status of the *HMGCR* promoter, we measured the extent to which template

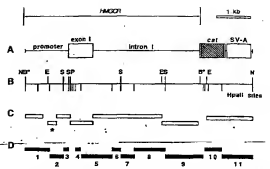


Fig. 1. Structure and analysis of the *HMGCR-cat* fusion transgene. (A) Structure of the injected construct *HMG1-cat*. The first *HMGCR* exon (noncoding) and the untranslated 5′ portion of the *HMGCR* second exon are indicated by open boxes; *cat*, hatched; the DNA sequence SV-A downstream from *cat* contains SV40 splice and polyadenylation signals derived from plasmid pSV2-CAT (Gorman et al., 1982). (B) Transgene restriction map. *Above*, restriction sites are B, *Bam*HI; B*, a *Bam*HI site destroyed during the cloning procedure; P, *Pvu*II, E, *Eco*RI, S, *Sma*I; flanking *Not*I sites (N) used for excision of the DNA construct prior to microinjection into fertilized mouse eggs are derived from the plasmid vector. *Below*, *Hpa*II/*Msp*I sites; the group of sites beneath the *Pvu*II site (P) is a cluster of seven sites over a region of 225 bp (spacing: 20, 25, 80, 30, 50 and 20 nt; unpublished data). (C) Hybridization probes employed to determine methylation status; the probe segment employed in the Southern analysis of Fig. 2 is marked with an asterisk. (D) DNA segments analysed for methylation inhibition of excision in Fig. 3.

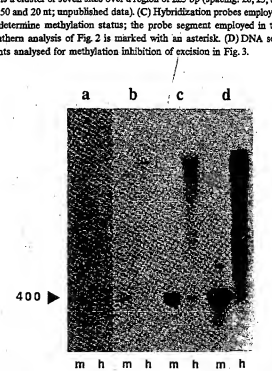


Fig. 2. Representative Southern analysis of *Hpa*II/*Msp*I digests. Liver DNA was prepared by a standard proteinase K/SDS/phenol procedure; 10 μg aliquots were digested with excess *Msp*I (m) or *Hpa*II (h), resolved by 1% agarose gel electrophoresis, and subjected to Southern analysis using a radiolabelled probe (marked with an asterisk in Fig. 1C) designed to detect the 400-bp segment 2 (arrowed) in Fig. 1D. Animals were: (a) wild-type; (b) transgenic line 74; (c) line 40; (d) line 78. The high-M_r bands of cleavage-resistant DNA are above the highest M_r marker employed (21.5 kb; not shown).

modification was able to inhibit excision of different transgene DNA segments from the genome. Total liver DNA from representative animals of different transgenic lines was cleaved either with *Hpa*II or with the methylation-insensitive isoschizomer *Msp*I, and after gel electrophoresis and Southern blotting examined for hybridization to separate probes (Fig. 1C) covering different regions of the *HMGCR* gene. The extent of methylation within each segment (Fig. 1D) of the transgene was assessed by the ratio of band intensities in the *Hpa*II and *Msp*I lanes. A typical experimental result is presented in Fig. 2 (zone analysed: segment 2 in Fig. 1D); data obtained for segments covering the entire transgene are compiled in Fig. 3.

Whereas excision of segments within the endogenous *HMGCR* promoter (two copies per diploid genome) in normal mice was not detectably blocked by methylation (<2%), flanking regions were resistant to digestion with *Hpa*II (Fig. 3a). In transgenic mice carrying ten copies of the *HMG1-cat* construct the *HMGCR* transgene promoter remained demethylated although the region devoid of methylation was narrower than that of the *HMGCR* gene in normal mice (Fig. 3b). However, methylation of the transgene promoter was observed to increase progressively in animals carrying 22 (35% methylation-inhibition of seg-

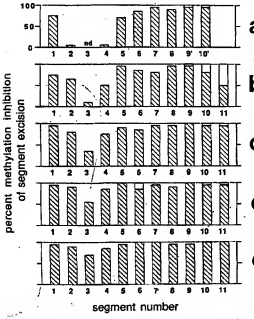


Fig. 3. Compilation of methylation data. Representative animals carrying different transgene copy numbers were: (a) wild type (2 copies of the *HMGCR* promoter); (b) transgenic line 74 carrying ten copies of the *HMG1-cat* fusion transgene); (c) line 40, 22 copies; (d) line 80, 45 copies; (e) line 78, 260 copies. Vertical axis: the extent of methylation inhibition of segment excision, determined as the ratio of the intensities of the relevant bands in the *Hpa*II and *Msp*I lanes of liver DNA (e.g., Fig. 2). Unshaded areas above certain histogram bars indicate 'greater-than' values. Horizontal axis: segment number (see Fig. 1D); internally-labelled probes used to determine methylation status (Mehtali, 1988) are presented in Fig. 1C. Quantitative scanning densitometry of autoradiograms (GS300 Scanner, Hoeffer Scientific) was used to determine the *Hpa*II/*Msp*I excision ratio. Segment 1 comprises two equally-sized *Hpa*II fragments that were not resolved by gel electrophoresis; segment 10 overlaps an *Hpa*II 'slow' site within *cat* that is partially refractory to *Hpa*II cleavage. Segments 9 and 10 represent the next two *Hpa*II fragments present in the endogenous *HMGCR* gene but not in the *HMG1-cat* transgene. For technical reasons the excision ratio for segment a3 wes not determined (*nd*).

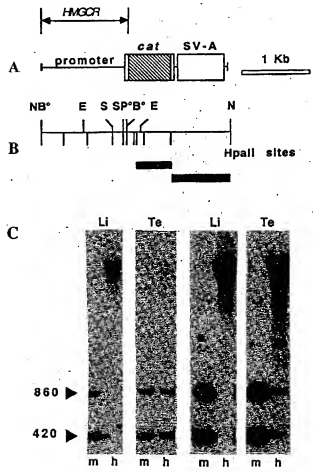


Fig. 4. Methylation state of the *HMG2-cat* transgene. (A) Structure of *HMG2-cat*: this construct is a derivative of *HMG1-cat*, in which the region between the *Pvu*II site in the *HMGCR* first exon and the site marked B* at the 5' end of *cat* in *HMG1-cat* (see Fig. 1A,B) has been removed. (B) Restriction site nomenclature is as in Fig. 1B; the solid bars given below indicate the segments detected by the hybridization probe employed (DNA fragment comprising *cat* and the SV40 splice and poly-adenylation signals, not shown). (Panels C) Southern hybridization of *Msp*I (m) and *Hpa*II (h) digests of total liver (Li) and testis (Te) DNA from representative animals bearing eleven (line 61) or 170 copies (line 76) of the fusion transgene. *Hpa*II fragments detected (sizes in bp) are arrowed. High-M_r band sizes and all experimental conditions are as in Fig. 2 legend.

TABLE I

CAT activity in tissues of *HMGCR-cat* transgenic mice

Construct[a]	Line	Copy No.[b]	CAT activity[d] in mouse tissues[c]						
			ta	li	in	ki	br	st	te
HMG1-cat	74[e]	10	12	5	21	5	62	31	59
HMG1-cat	40[e]	22	17	20	107	5	90	72	na
HMG1-cat	40	22	30	90	95	15	80	110	140
HMG1-cat	80[e]	45	6	32	21	7	70	23	na
HMG1-cat	78[e]	260	0.6	0.5	2	0.6	0.6	0.5	32
HMG2-cat	61[f]	11	0.6	0.9	0.7	0.6	0.7	0.8	135
HMG2-cat	76[f]	170	0.5	1.6	0.7	0.6	0.7	0.8	208

[a] Construct p*HMG1-cat* contains a 5.5-kb *Bam*HI segment of the mouse *HMGCR* gene (Mehtali, 1988) comprising the mouse *HMGCR* promoter region, the first (untranslated) exon, and the first intron, linked to the bacterial *cat* gene (Fig. 1A) and propagated in plasmid pPolyIII-I (Lathe et al., 1987). The downstream *Bam*HI site used in this construction lies immediately beyond the *HMGCR* exon II splice acceptor site (Mehtali, 1988; Gautier et al., 1989; M.M. and R.L., in preparation). In p*HMG2-cat* the region between the *Pvu*II site in the *HMGCR* first exon and the *BamHI*[e] site at the 5′ end of *cat* (see Fig. 1A) was removed: the structure of the fusion gene is presented in Fig. 4. Plasmid DNAs were propagated on *E. coli* 1106 (*thr leu thi hsdS*). *HMG-cat* fusion genes were excised from the vector by *Not*I digestion and purified by sucrose gradient centrifugation (10–30% sucrose, 35000 rpm, Beckman SW41 rotor, 14 h, 20°C). Two hundred copies were injected into fertilized eggs (C57B1/6 × SJL·F1 hybrids) and the presence of the transgene was detected by Southern blot analysis of DNA from the tails of four-week-old animals (Palmiter et al., 1982). Transgenic lines were established by systematic back-crossing with C57B1/6 × SJL·F1 hybrids.

[b] Copy numbers were determined by quantitative densitometry of Southern blots; all transgenic animals carry the transgene as an unrearranged tandem repeat integrated at a single autosomal location (data not presented).

[c] Tissues were: ta, tail; li, liver; in, intestine; ki, kidney; br, brain; st, stomach; te, testis.

[d] CAT activity in tissue homogenates (ultra-turrax, in 0.25 M Tris-HCl pH 7.8, 1 mM phenylmethylsulfonylfluoride) was determined, after heat treatment (10 min, 65°C) and clarification, by the transfer of radioactivity from ^{14}C acetyl CoA to ethyl-acetate-soluble Cm in a standard assay (Sleigh, 1986) at a constant protein concentration as measured using a commercial assay kit (BioRad), and is expressed as the incorporation of ^{14}C acetyl into Cm (cpm × 10^3) per 100 μg protein. The background value for CAT activity in negative extracts was 0.5. na, not applicable.

[e] Animals used for analysis in Fig. 2 (founder transgenic male of line 78; founder transgenic females of lines 40 and 80, and a second generation male for line 74; the remaining line 40 animal was a second generation male; M.M. and R.L., in preparation).

[f] Animals used for analysis were second generation males.

ment excision, Fig. 3c), 45 (48%) or 260 copies (78%) (Fig. 3d,e). The methylation pattern appeared identical in other tissues examined (intestine, testis) though the extent of the hypomethylated region was slightly larger in testis than in the other tissues (data not presented). The strength of the hybridization signal originating from the transgene unfortunately precluded examination of the methylation pattern of the endogenous *HMGCR* gene in these transgenic mice.

(c) *HMG2-cat* transgenic mice

To determine whether there is a general correlation between increased transgene copy number and loss of methylation-free status we constructed a deletion derivative of the *HMG1-cat* construct, *HMG2-cat*, in which a sub-fragment of the *HMGCR* promoter region is linked directly to *cat* (Fig. 4A). *HMG2-cat* was introduced into the mouse germline and tissues from transgenic animals were analysed for CAT activity (Table I). In the two lines examined (61 and 67; Table I) activity was detected in testis but not in other tissues. The methylation state of the fusion transgene in lines 61 and 76 was examined by *Hpa*II or *Msp*I digestion and Southern hybridization to probes designed to detect the reporter gene segment of the transgene (Fig. 4B) or the *HMGCR* promoter region (data not presented).

In liver (Fig. 4C) and intestine (data not presented), tissues in which no expression is observed (Table I), the fusion transgene was extensively methylated in both transgenic lines. In testis of mice harboring eleven copies of the transgene (line 61), a tissue in which the transgene is expressed, the *HMG2-cat* promoter and surrounding regions were essentially devoid of methylation (Fig. 4C; also data not presented). In contrast, the transgene in testis of line 76 (harboring 170 copies) was found to be substantially methylated (Fig. 4C), supporting a link between increased copy number and loss of methylation-free status. Nevertheless, in testis of line 76 a small proportion of transgene copies appeared to be devoid of methylation; these few unmethylated copies may be responsible for the observed transgene expression level. Indistinguishable methylation patterns were obtained in all cases irrespective of whether the probe employed covered the reporter gene segment (Fig. 4C) or the *HMGCR* promoter region (data not presented). It is of note that hypomethylation of the

HMGCR promoter region in testis of line 61 now extends into the adjacent bacterial *cat* gene (Fig. 4). Similar hypomethylation of *cat* was observed in testis of a further transgenic mouse line (No 60) harboring 15 copies of the *HMG2 -cat* transgene (data not presented).

(d) Conclusions

HTF islands are associated with control regions of active genes, particularly the housekeeping genes (Bird, 1986; Lavia et al., 1987; Gardiner-Garden and Frommer, 1987). It was previously reported that the methylation-free *HTF* island of the *Thy-1* gene is maintained when the intact gene is introduced into the mouse germline (Kolsto et al., 1986). We describe here that hypomethylation of a housekeeping gene (*HMGCR*) promoter is dependent upon transgene copy number, and methylation-free status of the *HTF* island at the 5′ end of a *HMGCR-cat* fusion transgene, *HMG1-cat*, is lost at increasing copy-number. In further transgenic mice bearing the *HMG2-cat* deletion derivative of the fusion construct, CAT activity and transgene hypomethylation were only detected in testis. It is of note that ectopic expression of tissue-specific transgenes in testis has been observed previously (Lacy et al., 1983; Shani, 1986; Al-Shawi et al., 1988; and our unpublished data). As observed in *HMGCR -cat* transgenic animals, methylation-free status in testis of *HMG2-cat* transgenic animals was also lost at high copy number.

One possible explanation for loss of hypomethylation at high copy number is out-titration of regulatory proteins or other binding factors that protect the DNA from methylation. Our data do not exclude the possibility that out-titration of a demethylase activity (e.g., Razin et al., 1986) might also reduce methylation level. However, methylation of the transgene promoter at high copy number demonstrates that *HTF* island DNA is not intrinsically resistant, in vivo, to methylation of C residues within CpG doublets. Because saturable factors thus appear to protect DNA from methylation in vivo, factor binding seems likely to precede methylation. In consequence, it would appear unlikely that the methylation status of the *HMGCR* promoter itself plays a major role in determining the extent of factor binding in vivo.

Despite the fact that the *HMGCR* promoter present on the transgene can outnumber the endogenous *HMGCR* promoter by a factor of 100, we have not detected any alteration in the expression of the endogenous gene (M.M., R.L. and G. Boukamel, unpublished data), in agreement with the conclusions of a study (Davis and MacDonald, 1988) using a rat elastase I transgene. Because high transgene copy number appears to be without effect on the expression of the endogenous *HMGCR* gene, binding factors may be only locally out-titrated, possibly arguing for restricted diffusion of binding factors (see Richetti et al., 1988).

Although we report saturable hypomethylation of the *HMGCR* promoter, this result may contrast with the report of Kolsto et al. (1986) who describe hypomethylation of a hybrid *Thy-1* gene promoter in mice bearing 60 copies of the transgene. However, the presence of extraneous sequences in *Thy-1* transgenic mice (Grosveld and Kollias, 1988) complicates interpretation. It therefore remains unclear whether loss of methylation-free status and/or local out-titration of binding factors is a general feature of large transgene arrays.

We also report that the hypomethylated region of the *HMGCR-HTF* island can extend into adjacent bacterial DNA (*cat*). This phenomenon was only observed when the *cat* reporter gene was linked directly to the *HMGCR* promoter region (construct *HMG2-cat*) and not when the *cat* segment was separated from the *HMGCR* regulatory region by 4 kb of intervening *HMGCR* genomic DNA (construct *HMG1-cat*). Hypomethylation of adjacent bacterial DNA was only observed in testis, the only tissue in which transgene expression was detected, and we speculate that cooperative binding of factors (e.g., Phillips et al., 1989; see also Murray and Grosveld, 1987) to the *HMGCR* promoter and to adjacent CpG-rich bacterial DNA may be responsible. We cannot however exclude the possibility that transcription per se can contribute to under-methylation.

Taken together, our data argue that the cytosine-methylase passively methylates DNA according to its accessibility/affinity for the methylase. Because DNA methylation can inhibit gene activity, passive methylation of transcriptionally inactive regions may contribute to the repressed state. Lower eukaryotes and invertebrates lack detectable DNA methylation, and the large genome sizes of vertebrates and plants may provide a selective advantage for DNA methylation (Antequera and Bird, 1988), for instance by marking untranscribed and passively-methylated DNA for higher-order condensation.

ACKNOWLEDGEMENTS

We would like to thank C. Marfing for skillful microinjection and P. Meyer and I. Kuhn for technical assistance. C. Tomasetto, J.O. Bishop, A.J. Clark, A. Pavirani, P. Gerlinger and P. Chambon are thanked for helpful discussions. This work was supported by grants from the CNRS, INSERM, Centre de Lutte contre le Cancer, ARC and the FRMF.

REFERENCES

Al-Shawi, R., Burke, J., Jones, C.T., Simons, J.P. and Bishop, J.O.: A MUP promoter-thymidine kinase reporter gene shows relaxed tissue-specific expression and confers male sterility upon transgenic mice. Mol. Cell. Biol. 8 (1988) 4821–4828.

Antequera, F. and Bird, A.P.: Unmethylated CpG islands associated with genes in higher plant DNA. EMBO J. 7 (1988) 2295–2299.

Bird, A.P.: DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res. 8 (1980) 1499–1504.

Bird, A.P.: CpG islands and the function of DNA methylation. Nature 321 (1986) 209–213.

Bird, A.P.: CpG islands as gene markers in the vertebrate nucleus. Trends Genet. 3 (1987) 342–347.

Brown, M.S. and Goldstein, J.L.: Multivalent feedback regulation of HMGCoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth. J. Lipid Res. 21 (1980) 505–517.

Cedar, H.: DNA methylation and gene activity. Cell 53 (1988) 3–4.

Cooper, D.N., Taggart, M.H. and Bird, A.P.: Unmethylated domains in vertebrate DNA. Nucleic Acids Res. 11 (1983) 647–658.

Davis, B.P. and MacDonald, R.J.: Limited transcription of rat elastase I transgene repeats in transgenic mice. Genes Dev. 2 (1988) 13–22.

Dynan, W.S.: Promoters for housekeeping genes. Trends Genet. 2 (1986) 196–197.

Gardiner-Garden, M. and Frommer, M.: CpG islands in vertebrate genomes. J. Mol. Biol. 196 (1987) 261–282.

Gautier, C., Mehtali, M. and Lathe, R.: A ubiquitous mammalian expression vector, pEHMG, based on a housekeeping gene promoter. Nucleic Acids Res. 17 (1989) 8389.

Gorman, C.M., Moffat, L.F. and Howard, B.H.: Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. Mol. Cell. Biol. 2 (1982) 1044–1051.

Grippo, F., Daccararino, M., Parasie, E. and Scarano, E.: Methylation of DNA in developing sea urchin embryos. J. Mol. Biol. 36 (1968) 195–208.

Grosveld, F. and Kollias, G.: Ectopic expression of Thy-1 in the kidneys of transgenic mice induces functional and proliferative abnormalities. Cell 54 (1988) 920.

Kolsto, A.-B., Kollias, G., Giguere, V., Isobe, K.-I., Prydz, H. and Grosveld, F.: The maintenance of methylation-free islands in transgenic mice. Nucleic Acids Res. 14 (1986) 9667–9678.

Lacy, E., Roberts, S., Evans, E.P., Burtenshaw, M.D. and Costantini, F.D.: Foreign beta-globin gene in transgenic mice: integration at abnormal chromosomal positions and in inappropriate tissues. Cell 34 (1983) 353–358.

Lathe, R., Vilotte, J.L. and Clark, A.J.: Plasmid and bacteriophage vectors for excision of intact inserts. Gene 57 (1987) 193–201.

Lavia, P., MacLeod, D. and Bird, A.: Coincident start sites for divergent transcripts at a randomly selected CpG rich island of mouse. EMBO J. 6 (1987) 2773–2779.

Luskey, K.L.: Structure and expression of 3-hydroxy 3-methylglutaryl coenzyme A reductase. Ann. N.Y. Acad. Sci. 478 (1986) 249–254.

Luskey, K.L.: Conservation of promoter sequence but not complex intron-splicing pattern in human and hamster genes for 3-hydroxy 3-methylglutaryl coenzyme A reductase. Mol. Cell. Biol. 7 (1987) 1881–1893.

Mehtali, M.: Ph.D. thesis, University of Strasbourg, 1988.

Murray, E.J. and Grosveld, F.: Site-specific demethylation in the promoter of human gamma-globin gene does not alleviate methylation-mediated suppression. EMBO J. 6 (1987) 2329–2335.

Palmiter, R.D. and Brinster, R.L.: Germ line transformation of mice. Annu. Rev. Genet. 20 (1986) 465–499.

Palmiter, R.D., Chen, H.Y. and Brinster, R.L.: Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring. Cell 29 (1982) 701–710.

Phillips, S.E.V., Manfield, I., Parsons, I., Davidson, B.E., Rafferty, J.B., Somers, W.S., Margarita, D., Cohen, G.N., Saint-Girons, I. and Stockley, P.G.: Cooperative tandem binding of *met* repressor of *Escherichia coli*. Nature 341 (1989) 711–715.

Razin, A., Szyf, M., Kafri, T., Roll, M., Giloh, H., Scarpa, S., Carotti, D. and Cantoni, G.L.: Replacement of 5-methylcytosine by cytosine: a possible mechanism for transient DNA methylation during differentiation. Proc. Natl. Acad. Sci. USA 83 (1986) 2827–2831.

Reynolds, G.A., Basu, S.K., Osborne, T.F., Chin, D.J., Gil, G. and Luskey, K.L.: HMG CoA reductase: a negatively-regulated gene with unusual promoter and 5′ untranslated regions. Mol. Cell. Biol. 7 (1985) 1881–1893.

Richetti, M., Metzger, W. and Heumann, H.: One-dimensional diffusion of *Escherichia coli* DNA-dependent RNA polymerase: a mechanism to facilitate promoter location. Proc. Natl. Acad. Sci. USA 85 (1988) 4610–4614.

Shani, M.: Tissue-specific and developmentally-regulated expression of a chimeric actin-globin gene in transgenic mice. Mol. Cell. Biol. 6 (1986) 2624–2631.

Sleigh, M.: A non-chromatographic assay for expression of the chloramphenicol-acetyl-transferase gene in eucaryotic cells. Anal. Biochem. 156 (1986) 257–262.

TAB 5

This material may be protected by Copyright law (Title 17 U.S. Code)

Cell, Vol. 32, 1105–1113, April 1983, Copyright © 1983 by MIT

0092-8674/83/041105-09 $02.00/0

Independent Mechanisms Involved in Suppression of the Moloney Leukemia Virus Genome during Differentiation of Murine Teratocarcinoma Cells

Ohtsura Niwa* Yoshifumi Yokota,*
Hiroyuki Ishida* and Tsutomu Sugahara†
* Department of Experimental Radiology
Faculty of Medicine
Kyoto University
Sakyo-ku, Kyoto 606, Japan
† Kyoto National Hospital
Fushimi-ku, Kyoto 612, Japan

Summary

Expression and DNA methylation of the Moloney murine leukemia virus (M-MuLV) genome were investigated in murine teratocarcinoma cells after virus infection. The newly acquired viral genome was devoid of methylation, yet its expression was repressed. The integrated viral genome in undifferentiated teratocarcinoma cells was methylated within 15 days after infection. Although 5-azacytidine decreased the level of DNA methylation, it did not activate M-MuLV in undifferentiated cells. Activation by 5-azacytidine occurred only in differentiated teratocarcinoma cells. Thus two independent mechanisms seem to regulate gene expression during the course of differentiation. The first mechanism operates in undifferentiated cells to block expression of M-MuLV and other exogeneously acquired viral genes, such as SV40 and polyoma virus, and does not depend on DNA methylation. The second mechanism relates only to differentiated cells and represses expression of genes in which DNA is methylated.

Introduction

Laboratory strains of mice as well as field isolates of Mus species carry a large number of endogenous type C virus genomes. Blot hybridization analysis of mouse DNA digested with various restriction enzymes suggested that cells of inbred mice have as many as 50 copies of DNA sequences related to type C viruses (Dolberg et al., 1981). Inducible murine endogenous viruses are classified into two host range groups, ecotropic viruses and xenotropic viruses. Chromosomal locations of ecotropic viruses can be determined by classical Mendelian genetics. Ecotropic virus loci of AKR mice, *Akv1* and *Akv2*, have been mapped on chromosome 7 (Rowe et al., 1972) and on chromosome 16 (Kozak and Rowe, 1980), respectively. The *Cv* locus, an ecotropic virus locus of C3H/HeJ and BALB/c strains was detected on chromosome 5 (Kozak and Rowe, 1979; Ihle et al., 1979). C3H/Fg mice have an ecotropic virus locus on chromosome 7 at a site distinct from the *Akv1* locus. C58 and B10·BR mice carry unnamed ecotropic virus loci on chromosome 8 and on chromosome 11, respectively. Thus the distribution of endogenous ecotropic virus loci over a wide range of mouse chromosomes suggests that these endogenous virus genomes may have recently derived from a common prototype virus through germline integration during establishment of laboratory mice. Actually, Jaenisch demonstrated that infection of preimplantation stage BALB/129 mouse embryos with M-MuLV leads to integration of the virus into the germ line, in an endogenous unexpressed state (Jaenisch, 1976).

Infection of mouse embryo fibroblasts with murine ecotropic viruses usually results in stable integration of the viral genome and subsequent production of progeny virus. This makes for a clear distinction between fibroblast lines and the cells of preimplantation stage embryos where infection with exogenous viruses results in repression of the integrated virus genomes. Murine teratocarcinoma cells are the malignant and pluripotent stem cells derived from carcinoma of preimplantation stage embryos (Lehman et al., 1974). Teratocarcinoma stem cells infected with ecotropic murine type C viruses do not produce progeny virus (Peries et al., 1977; Teich et al., 1977; Huebner et al., 1979; Gautsch, 1980). The infecting viral genome, however, is carried in these cells and can be activated with bromodeoxyuridine (BrdUrd) (Speers et al., 1980).

The expression of endogenous virus genomes in avian and mouse embryo fibroblasts is regulated by DNA modification and can be activated by treating the cells with 5-azaCytidine (5-AzaCyd), a potent inhibitor of DNA methylation (Groudine et al., 1981; Niwa and Sugahara, 1981). We studied the state of the M-MuLV genome in teratocarcinoma cells and found that the viral genome in undifferentiated stem cells is repressed not by DNA methylation but by some other mechanism. Only after differentiation of the cells is the viral sequence under the control of a mechanism regulated by DNA modification.

Results

Repression of M-MuLV in Undifferentiated Teratocarcinoma Cells, and Induction of Its Expression by BrdUrd and Retinoic Acid

EC-A1 cells derived from PCC4 cells, a pluripotent stem cell line of mouse teratocarcinoma cells, were infected with M-MuLV. Cells were treated for 24 hr with BrdUrd and/or retinoic acid, a compound that induces differentiation of teratocarcinoma cells (Strickland and Mahdavi, 1978) before or after virus infection. Since both BrdUrd and retinoic acid exerted potent cytotoxic effects on EC-A1 cells, the virus-producing cells in the drug-treated cultures may be preferentially eliminated to yield a false negative result. Thus EC-A1 cells were treated with BrdUrd and/or retinoic acid before or after virus infection, and the cells were then cocultivated with SC-1 cells. The SC-1 cells were passaged once and assayed for virus

expression. The results in Table 1 indicate that preinfection and postinfection treatment with BrdUrd alone was effective for productive infection of the virus. Preinfection treatment with retinoic acid made the EC-A1 cells susceptible to infection with M-MuLV. The expression of the virus in the BrdUrd-treated cells was enhanced by the presence of retinoic acid. Although less efficient, postinfection treatment of the cells with BrdUrd together with retinoic acid did induce expression of the virus. A low but definite production of M-MuLV was detected by amplification through SC-1 cells when the EC-A1 cells were treated for 24 hr with 0.3–1 M of retinoic acid alone, immediately after infection.

Infection of F-9 cells, a nullipotent teratocarcinoma cell line (Bernstine et al., 1973), with M-MuLV also required preinfection or postinfection treatment with BrdUrd for efficient expression of the virus (O. Niwa, unpublished observation).

The teratocarcinoma stem cells, EC-A1 and F-9 cells, were all derived from 129 mice that contained the ecotropic virus sequences and were mostly subgenomic in size (Chan et al., 1980). These have been classified as a no virus strain (Chattopadhyay et al., 1974). The virus which was recovered from BrdUrd-treated EC-A1 cells, and which had been infected with M-MuLV, grew equally well on NIH/3T3 and BALB/3T3 cells. The tropism of the virus and the lack of virus activation from uninfected EC-A1 cells (O. Niwa, unpublished observation) strongly suggest that the virus thus recovered from infected EC-A1 cells after BrdUrd treatment is M-MuLV and not the virus endogenous to EC-A1 cells.

Table 1. Plaque-Forming Cells per 2×10^5 SC-1 Cells Cocultivated with EC-A1 Cells Treated Preinfection and Postinfection with BrdUrd and/or Retinoic Acid

BrdUrd (μg/ml)	Retinoic Acid (μM)	Preinfection Treatment/ Postinfection Treatment
0	0	0/0
0	0.03	5/0
0	0.1	14/0
0	0.3	23/5
0	1.0	46/3
5	0	122/7
5	0.03	634/7
5	0.1	TMTC/25
5	0.3	TMTC/31
5	1.0	TMTC/23
10	0	TMTC/6
10	0.03	TMTC/15
10	0.1	TMTC/43
10	0.3	TMTC/22
10	1.0	TMTC/19
20	0	TMTC/12
20	0.03	TMTC/21
20	0.1	TMTC/48
20	0.3	TMTC/39
20	1.0	TMTC/51

TMTC: too numerous to count.

Isolation of EC-A1(Mo) Clones Carrying M-MuLV Genome in a Repressed State

EC-A1 cells were infected with M-MuLV at a moi of 1.0, which had been determined by titration on SC-1 cells. The cultures were then trypsinized, and the cells were plated for ring cloning. Each of the randomly isolated clones was treated with 20 μg/ml of BrdUrd and cocultivated with SC-1 cells. SC-1 cells were passaged three times and tested for virus expression by the reverse XC test. Of 110 clones thus tested, 47 expressed virus after BrdUrd treatment. These virus-inducible clones, designated EC-A1(Mo) clones, therefore carry the M-MuLV genome, in a repressed form.

Transcriptional Control of M-MuLV Expression in Undifferentiated Teratocarcinoma Cells

Liquid hybridization experiments indicated no detectable viral RNA transcript in M-MuLV-infected teratocarcinoma cells (Teich et al., 1977). We have isolated total cellular RNA from one of the virus-carrying clones, EC-A1(Mo)4. RNA was size-differentiated on agarose gel electrophoresis, transferred to diazobenzyloxymethyl paper (DBM paper) and analyzed for the sequence hybridizable to ^{32}P-labeled M-MuLV DNA (Figure 1). SC-1 cells productively infected with M-MuLV had two major bands, 34S and 24S, corresponding to virus genomic RNA and spliced *env* gene messenger RNA. On the other hand, uninfected SC-1 cells, EC-A1 cells and EC-A1(Mo)4 cells contained no detectable level of RNA sequence hybridizable to M-MuLV DNA. The same filter was hybridized with an 18S rRNA probe. Three bands corresponding to 18S rRNA and two precursor RNAs were detected in all four RNA samples. The density of the bands revealed by this probe varied little among the four cell lines tested, indicating that the amounts of RNA used for the analysis were the same for these cell lines. Total cellular RNA isolated from EC-A1 cells 48 hr after infection with M-MuLV was analyzed by the dot blot hybridization technique (data not shown). Here too, the RNA transcript of M-MuLV was not detected.

5-AzaCyd Induction of M-MuLV Gene Expression and the State of Differentiation of Teratocarcinoma Cells

We attempted to induce M-MuLV expression in undifferentiated EC-A1(Mo) clones by 5-AzaCyd. To our surprise, none of the 47 EC-A1(Mo) clones expressed virus after treatment with 2 μg/ml 5-AzaCyd, although virus was readily recovered when the clones were treated with BrdUrd.

To determine whether or not the DNA methylation was suppressed by 5-AzaCyd in undifferentiated teratocarcinoma cells, two clones, EC-A1(Mo)1 and EC-A1(Mo)4, were treated with 5-AzaCyd, and the level of DNA methylation was measured. As is clear from

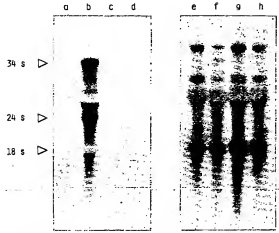


Figure 1. Blot Hybridization of Total Cellular RNA Isolated from Four Cell Lines

Probes used were ^{32}P-labeled M-MuLV DNA (lanes a, b, c and d) and ^{32}P-labeled cloned rRNA gene (lanes e, f, g and h). (Lanes a and e) RNA from SC-1 cells; (lanes b and f) RNA from SC-1 cells infected with M-MuLV; (lanes c and g) RNA from EC-A1 cells; (lanes d and h) RNA from EC-A1(Mo)4 cells.

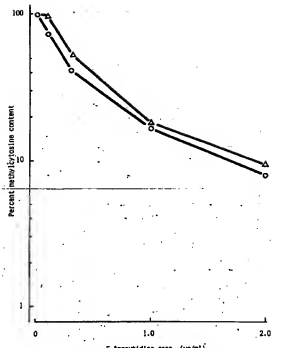


Figure 2. Level of Methylcytosine in Newly Replicated DNA of EC-A1 Cells

Cells were seeded onto 3 cm dishes at a concentration of 1 × 10^6 cells per dish. After overnight incubation, the dishes were nourished with medium containing (methyl-^{3}H)methionine at 4 μCi/ml and (2-^{14}C)thymidine at 0.005 μCi/ml. DNA was collected, hydrolyzed and separated on a cellulose thin-layer glass plate. The ratio of ^{3}H and ^{14}C counts in methylcytosine and thymidine respectively was taken as a relative measure of methylcytosine content (○) EC-A1(Mo)1 cells; (△) EC-A1(Mo)4 cells.

Figure 2, the level of methylcytosine in cellular DNA decreased with increasing doses of 5-AzaCyd. Therefore, in undifferentiated teratocarcinoma cells, suppression of DNA methylation does not seem to result in activation of the viral genome.

EC-A1(Mo)4 cells were treated with dimethylacetamide for 2 weeks to induce differentiation, and such was assessed to be complete by the morphology. M-MuLV was never expressed in the differentiated EC-A1(Mo)4 cells. These cells were then treated with 5-AzaCyd and cocultivated with SC-1 cells. The SC-1 cells were passaged twice, and the reverse XC test was performed at each passage. Only in the differentiated EC-A1(Mo)4 cells was the 5-AzaCyd-activated expression of M-MuLV evident (Table 2). Therefore, at least in the differentiated EC-A1(Mo)4 cells, the viral genome seems to be regulated by the extent of DNA methylation, and suppression of DNA methylation activates expression of the virus.

Table 2. Plaque-Forming Cells per 2 × 10^6 SC-1 Cells Cocultivated with EC-A1(Mo)4 Cells Treated with 5-AzaCyd

Cells	Concentration of 5-AzaCyd (μg/ml)	First Passage/ Second Passage
EC-A1(Mo)4 cells undifferentiated	2	0/0
	4	0/0
	6	0/0
	8	0/0
EC-A1(Mo)4 cells differentiated	1	0/0
	2	2/31
	4	13/TMTC

TMTC: too numerous to count.

Lack of DNA Methylation of Unintegrated M-MuLV Genome

The data presented above indicate that the M-MuLV genome in the differentiated teratocarcinoma cells is methylated. Experiments were then designed to determine the timing of DNA methylation of the M-MuLV genome, after virus infection of the undifferentiated cells.

EC-A1 cells were infected with M-MuLV, and DNA was extracted from the Hirt supernatant fraction which contained unintegrated viral genome. DNA isolated from EC-A1 cells 6 hr after infection contained three molecular species of M-MuLV—namely, closed circular, linear and open circular DNAs (Figure 3, lane a). Closed circular DNA and open circular DNA consisted of two subbands differing slightly in size. A linear molecule had a single band of 8.8 kb. Upon cleavage with Hind III, two bands corresponding to 8.2 kb and 8.8 kb linear molecules were detected (Figure 3, lane b) and were assumed to be the full-sized M-MuLV genome with one and two long terminal repeats (LTRs). All of these sequences were com-

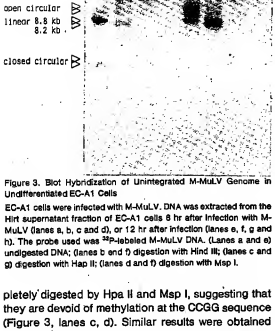


Figure 3. Blot Hybridization of Unintegrated M-MuLV Genome in Undifferentiated EC-A1 Cells

EC-A1 cells were infected with M-MuLV. DNA was extracted from the Hirt supernatant fraction of EC-A1 cells 8 hr after infection with M-MuLV (lanes a, b, c and d), or 12 hr after infection (lanes e, f, g and h). The probe used was ^{32}P-labeled M-MuLV DNA. (Lanes a and e) undigested DNA; (lanes b and f) digestion with Hind III; (lanes c and g) digestion with Hap II; (lanes d and f) digestion with Msp I.

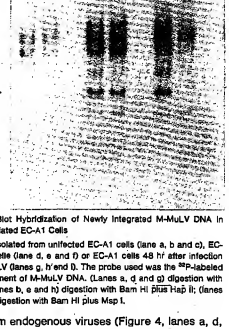


Figure 4. Blot Hybridization of Newly Integrated M-MuLV DNA in Undifferentiated EC-A1 Cells

DNA was isolated from unifected EC-A1 cells (lane a, b and c), EC-A1(Mo)4 cells (lane d, e and f) or EC-A1 cells 48 hr after infection with M-MuLV (lanes g, h and i). The probe used was the ^{32}P-labeled Sma I fragment of M-MuLV DNA. (Lanes a, d and g) digestion with Bam HI; (lanes b, e and h) digestion with Bam HI plus Hap II; (lanes c, f and i) digestion with Bam HI plus Msp I.

pletely digested by Hpa II and Msp I, suggesting that they are devoid of methylation at the CCGG sequence (Figure 3, lanes c, d). Similar results were obtained with DNA isolated at 12 hr after infection (Figure 3, lanes e, f, g, h). However, one difference in the 12 hr sample was that this DNA lacked a closed circular molecule (Figure 3, lane e). It is of interest that the Hirt supernatant DNA isolated 24 hr after infection had a much lesser amount of viral DNA, while the 48 hr sample was devoid of M-MuLV sequences (data not shown). In the case of SC-1 cells, DNA extracted 24 hr after infection had the greatest amount of unintegrated M-MuLV provirus sequence (O. Niwa, unpublished observation).

Undermethylation of Freshly Integrated M-MuLV Genome and Its Subsequent Methylation during Multiple Cell Cycling of Undifferentiated Cells

EC-AI cells were infected with M-MuLV at a moi of 2, and DNA was isolated 48 hr later. The integration of viral genome occurs at random sites. Digestion of DNA from randomly infected cells with Eco RI, which does not cleave the M-MuLV genome, will produce multiple fragments of various sizes carrying integrated viral sequences flanked by cellular sequences, and these cannot be resolved by agarose gel electrophoresis. Therefore, we used Bam HI, which cut the M-MuLV genome internally. As a probe to detect viral genome, the 316 base Sma I fragment of M-MuIV DNA was labeled with ^{32}P. This probe allows for detection of the 3 kb internal fragment of Bam HI–digested M-MuLV together with other sequences derived from endogenous viruses (Figure 4, lanes a, d, g). Uninfected EC-A1 cells lacked the 3 kb M-MuLV-specific sequence (Figure 4, lane a). DNA of EC-A1 cells 48 hr after infection contained the 3 kb band (Figure 4, lane g), and this fragment was not derived from the unintegrated M-MuLV, as unintegrated proviral DNA disappeared from the Hirt supernatant fraction of EC-A1 cells before 48 hr of infection. Double digestion of 48 hr postinfection DNA with Hap II and with Msp I completely eliminated the 3 kb Bam HI band, suggesting that this part of the M-MuLV genome is devoid of DNA methylation (Figure 4, lanes h, i). DNA was isolated from EC-A1(Mo)4 cells that had undergone multiple cell cycling after virus infection. The M-MuLV-specific 3 kb Bam HI band was clearly demonstrable in the cellular DNA (Figure 4, lane d). However, the same sequence was now resistant to Hap II digestion (Figure 4, lane e). Other bands derived from endogenous virus genomes were also resistant to Hap II (Figure 4, lanes b, e, h). DNAs from another EC-A1(Mo) clone, EC-A1(Mo)22, and randomly infected EC-A1 cells passaged at least 20 times also contained the Hpa II–resistant 3 kb Bam HI fragment specific to M-MuLV.

Experiments were performed to determine the time of the DNA methylation of M-MuLV genome after infection. DNA was isolated from M-MuLV-infected EC-A1 cells on days 4, 10 and 15 after infection, and analyzed for the state of DNA methylation (Figure 5). The M-MuLV genome was devoid of methylation at

Hap II sites up to day 10 (Figure 5, lanes b, e). However, the viral genome was resistant to Hap II when it was isolated from 15-day-old cultures. Therefore, the integrated M-MuLV genome was methylated between days 10 and 15 in culture. Production of M-MuLV from infected EC-A1 cells during these 10 days was nil.

These results indicate that although unintegrated and newly integrated M-MuLV genomes were devoid of DNA methylation, the same sequence is methylated in the undifferentiated cells kept in culture for over 15 days.

Lack of Change in the State of DNA Methylation of the M-MuLV Genome during Differentiation of Teratocarcinoma Cells

DNA was isolated from undifferentiated EC-A1(Mo)4 cells and dimethylacetamide-induced differentiated EC-A1(Mo)4 cells. These DNAs contained the 3 kb Bam HI fragment of the integrated M-MuLV genome, as well as other bands derived from endogenous virus genomes (Figure 6). The 3 kb fragment of M-MuLV in undifferentiated EC-A1(Mo)4 cells was again resistant to digestion with Hap II, confirming the result in Figure 4, lane e (Figure 6, lane b). Similar resistance to Hap II enzyme was noted when the DNA from differentiated EC-A1(Mo)4 cells was analyzed. Therefore, the pattern of DNA methylation of at least the 3 kb Bam HI fragment did not change during differentiation of the cells, yet inducibility of the viral genome by 5-AzaCyd treatment changed drastically.

Transfection with DNA from M-MuLV Infected Teratocarcinoma Cells

DNA could be isolated from EC-A1 cells 2 days after infection with M-MuLV and also from EC-A1(Mo) clones. SC-1 cells were transfected with these DNAs

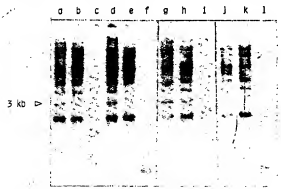


Figure 5. Blot Hybridization of Newly Integrated M-MuLV DNA in Undifferentiated EC-A1 Cells

DNA was isolated from EC-A1 cells on the day 4 (lanes a, b and c), day 10 (lanes d, e and f) or day 15 (lanes g, h and i) after infection with M-MuLV, or from uninfected EC-A1 cells (lanes j, k and l). (Lanes a, d, g and j) digestion with Bam HI; (lanes b, e, h and k) digestion with Bam HI plus Hap II; (lanes c, f, i and l) digestion with Bam HI plus Msp I.

(Table 3). Although at a low efficiency, DNA from EC-A1 cells infected with M-MuLV 2 days previously was capable of producing M-MuLV, while DNA from EC-A1(Mo) clones was not.

Discussion

Jaenisch and coworkers have demonstrated that endogenous virus can be formed by infection of the preimplantation stage embryo cells, with exogenous virus (Jaenisch et al., 1975; Jaenisch, 1976). Infection of undifferentiated stem cells of mouse teratocarcinoma leads to silencing of the exogenously acquired

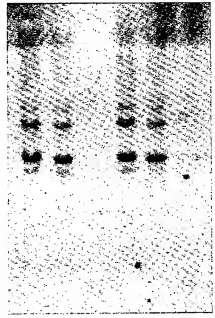


Figure 6. Blot Hybridization of M-MuLV Genome in Undifferentiated and Differentiated EC-A1(Mo)4 Cells

DNA was isolated from undifferentiated (lanes a, b and c) and differentiated EC-A1(Mo)4 cells (lanes d, e and f). The probe used was a ^{32}P-labeled 1.65 kb Bam HI–Hind III fragment of M-MuLV DNA. (Lanes a and d) digestion with Bam HI; (lanes b and e) digestion with Bam HI plus Hap II; (lanes c and f) digestion with Bam HI plus Msp I.

Table 3. Transfection of SC-1 Cells with DNA from M-MuLV-Infected Cells

DNA Source	DNA per Dish (μg)	Virus-Positive/ Total
SC-1	20	0/20
SC-1 M-MuLV-Infected	20	23/30
EC-A1	20	0/20
EC-A1(Mo)4	20	0/20
EC-A1(Mo)22	20	0/20
EC-A1(Mo)26	20	0/20
EC-A1 M-MuLV-infected*	20	9/20

* DNA was isolated 2 days after infection with M-MuLV.

type C virus genome (Teich et al., 1977; Speers et al., 1980). However, the mechanism of repression of integrated viral genome in the cells of the preimplantation stage embryo and in teratocarcinoma cells has remained unknown.

Undifferentiated teratocarcinoma cells have characteristics of preimplantation stage embryo cells in that they lack H-2 antigen on the cell surface (Artzt and Jacob, 1974), possess two active X chromosomes in female cells (Martin et al., 1978; McBurney and Strutt, 1980) and have the potential to differentiate into a variety of cell types (Kleinsmith and Pierce, 1964; Mintz and Illimensee, 1975). Thus teratocarcinoma cells are a pertinent model for studies on the undifferentiated state of embryogenesis.

For cells to be in an undifferentiated state, they must be equipped with the potential to suppress specifically expression of genes required only for differentiated cells. Mouse teratocarcinoma cells are resistant to exogenously incorporated genetic elements such as SV40 and polyoma virus (Swartzendruber and Lehman, 1975; Swartzendruber et al., 1977; Segal and Khoury, 1979) as well as to murine leukemia virus. These viral genomes may be regarded as unnecessary luxury genes in undifferentiated teratocarcinoma cells.

We found that the unintegrated M-MuLV genome and at least part of the newly integrated genome, the 3 kb Bam HI fragment, in undifferentiated teratocarcinoma cells are devoid of DNA methylation, yet the expression of the genome is repressed. SV40 DNA was also shown to be undermethylated in teratocarcinoma cells (Friedrich and Lehman, 1981). Treatment with retinoic acid, which triggers differentiation of the cells, can activate virus expression in undifferentiated cells, provided that drug treatment follows shortly after infection with M-MuLV, when methylation of the M-MuLV genome has not yet occurred (Table 1). However, the frequency of virus expression induced by retinoic acid was rather low, and for the entire course of differentiation about 10 days were required. The low frequency thus observed might be due to the incompleteness of differentiation. The finding that the freshly integrated M-MuLV genome in undifferentiated teratocarcinoma cells is transfectionally active on SC-1 cells (Table 3) suggests that this genome could be transcriptionally active in differentiated cells. Therefore, the M-MuLV genome freshly acquired by EC-A1 cells seems to be devoid of DNA modification, and DNA modification is known to reduce the rate of transcription (Stuhlmann et al., 1981; Hoffmann et al., 1982). Actually, the M-MuLV genome was devoid of methylation at Hpa II sites for at least 10 days after infection; it was not expressed in the undifferentiated cells. The 3 kb fragment of M-MuLV genome is subsequently methylated in infected cells kept in culture for over 15 days. The virus genome becomes transfectionally inactive and cannot be induced by simple differentiation of the cells. Also, suppression of DNA methylation by 5-AzaCyd did not activate expression of the virus genome in EC-A1(Mo) clones. Dot blot hybridization analysis of RNA isolated from 5-AzaCyd-treated EC-A1(Mo)4 cells that carry the methylated M-MuLV genome did not contain RNA hybridizable to M-MuLV probe, while in the BrdUrd-treated culture there was a marked increase of M-MuLV transcript (O. Niwa, unpublished observation). Therefore, we conclude that the M-MuLV genome is repressed in the undifferentiated cells by a mechanism other than DNA methylation. This repression can be unblocked by the treatment of the cells with BrdUrd. EC-A1 cells freshly infected with M-MuLV, and carrying the unmethylated viral genome, and EC-A1(Mo) clones presumably carrying the methylated M-MuLV genome were both induced by BrdUrd treatment (see Table 1 and above). Incorporation of BrdUrd into DNA does not affect the level of methylation (unpublished observations).

DNA methylation has no effect on the transcriptional activity of the M-MuLV genome in undifferentiated cells. The M-MuLV genome is nevertheless methylated in cells that have undergone multiple cell cycling. Genes that are not transcribed may be preferentially methylated in undifferentiated cells.

The pattern of DNA methylation of the 3 kb Bam HI fragment did not change during differentiation of EC-A1(Mo)4 cells. However, the M-MuLV genome is now susceptible to induction by 5-AzaCyd. Repression of the viral genome by DNA methylation thus seems operative only in differentiated cells. DNA methylation is thought to suppress gene expression through a condensation of chromatin, and antisera raised against methylcytosine bind to the heterochromatic region of the mouse chromosome (Miller et al., 1974). Methylcytosine is more abundant in the fraction of chromation that is resistant to DNAase digestion (Razin and Ceder, 1977). Methylated sequences of endogenous viruses are located on chromatin regions that are resistant to DNAase I (van der Putten et al., 1982).

Cells of the preimplantation stage embryo have two active X chromosomes and lack heterochromatin (Epstein et al., 1978). Heterochromatinization of an X chromosome occurs during differentiation of teratocarcinoma cells (Martin et al., 1978; McBurney and Strutt, 1980). Condensation of the heterochromatic region may even be facilitated by the presence of yet unidentified chromatin protein(s), and differentiation of cells may trigger production of this chromatin protein, which inactivates already methylated luxury genes by condensation of their chromatin regions.

In light of all these data, we propose the presence of two independent mechanisms regulating gene expression in mammalian cells. The first mechanism which is operating in undifferentiated cells, is not influenced by the state of DNA methylation for its function. Repression of transcription by this mecha-

nism may involve discrimination against the luxury gene promoter. Host range mutants of polyoma virus that can replicate on undifferentiated teratocarcinoma cells were found to possess mutations at the promoter region (Sekikawa and Levine, 1981). The second mechanism suppresses the expression of methylated genes by changing the conformation of their chromatin domains, and this mechanism operates only in differentiated cells. When M-MuLV infects undifferentiated teratocarcinoma cells, the viral genome is suppressed by the first mechanism. The transcriptionally inactive genome of M-MuLV may be gradually methylated during replication of the host cells. Differentiation of the cells terminates regulation by the first mechanism and activates the second mechanism, which now recognizes the methylated genome of M-MuLV and represses its expression by condensation of the chromatin domain. 5-AzaCyd unblocks the second mechanism by decreasing the level of methylcytosine, while BrdUrd unblocks both mechanisms. Since BrdUrd is known to change the affinity of DNA-binding proteins for DNA (Lin and Riggs, 1972, 1976; Goeddel et al., 1977), binding of some chromatin protein(s) to the promoter region of M-MuLV may be responsible for operation of the first mechanism. The first mechanism of the repression of gene expression seems to be *trans*-acting, since the extrachromosomal genome of SV40 and the M-MuLV genome integrated at random sites in undifferentiated cells are repressed. These findings suggest that the first mechanism may inactivate gene expression, possibly by some diffusible repressor-like protein, and if such is the case, the *trans*-acting nature can be readily explained.

During preparation of this manuscript, a report appeared on a similar subject (Stewart et al., 1982) These authors' data suggest that the M-MuLV genome becomes methylated immediately after integration. This difference between our results and theirs might relate to different cell lines. We used PCC4-derived EC-A1 cells, while they used F-9 cells. Gautsch and Wilson (1983) apparently obtained findings similar to those we report here.

Experimental Procedures

Cell Lines and Viruses

EC-A cells (Gautsch, 1980), a subline of PCC4 Azat cells (Jakob et al., 1973), were kindly provided by J. Gautsch. EC-A1 cells, a subclone of EC-A cells isolated in our laboratory, were also used. F-9 cells (Strickland and Mahdavi, 1978) were obtained from K. Sekigawa. SC-1 cells, a mouse embryo fibroblast line derived from a feral mouse (Hartley and Rowe, 1975), were obtained from A. Dedeve. Teratocarcinoma cell lines were grown in Eagle's minimal essential medium supplemented with 10% heat-inactivated fetal calf serum. SC-1 cells were grown in MEM alpha medium (Gibco) supplemented with 10% heat-inactivated calf serum.

M-MuLV was obtained from A. Ishimoto and grown on SC-1 cells.

Virus Infection, Transfection and Biological Assays

Procedures of virus infection were as described previously (Niwa et al., 1973). Titration of M-MuLV stock was on SC-1 cells by UV-XC assay (Rowe et al., 1970). M-MuLV-infected teratocarcinoma cells were assayed directly or after cocultivation with SC-1 cells by the reverse XC cell assay (Niwa et al., 1973).

Transfection of SC-1 cells with DNA from M-MuLV-infected EC-A1 cells was carried out by a procedure described by other workers (Stuhlmann et al., 1981).

Drug Treatments

Stock solutions of retinoid and BrdUrd were prepared in dimethylsulfoxide at concentrations of 10 mM and 1 mg/ml, respectively, and kept in the dark at −20°C. Teratocarcinoma cells were incubated with these drugs for 24 hr at 37°C either before or after infection with M-MuLV. When cocultivation with SC-1 cells was used for the amplification of progeny virus, teratocarcinoma cells were treated for 30 min with 25 μg/ml mitomycin C to suppress overgrowth of the undifferentiated cells.

Differentiated teratocarcinoma cells were obtained by the procedure of Speers et al. (1980) with slight modification. Undifferentiated teratocarcinoma cells grown as monolayer cultures were nourished every 3 days with a medium containing 10 mM dimethylacetamide for 2 weeks. Cultures consisted only of cells with an epithelial morphology. Although retinoic acid also induced differentiation of EC-A1 cells, it was more cytotoxic than dimethylacetamide.

Base Analysis

The level of methylcytosine in the newly synthesized DNA was assayed as described previously (Niwa and Sugahara, 1981).

Extraction of Cellular RNA and DNA

Total cellular RNA was isolated by sedimentation through cesium chloride as described by others (Chirgwin et al., 1979).

For the isolation of DNA, dishes were washed with phosphate-buffered saline solution and digested at 37°C for 2 hr with 100 μg/ml RNAase A in 1% sodium dodecylsulfate, 0.1 M NaCl, 5 mM EDTA and 20 mM Tris–HCl (pH 8.0). Proteinase K was then added at 100 μg/ml, and dishes were incubated for another 2 hr at 37°C. The lysate was extracted three times with phenol–chloroform and ethanol-precipitated. DNA thus extracted was used for further analysis. Unintegrated viral DNA was isolated by the procedure of Hirt (1967).

Restriction Endonuclease Digestion, Gel Electrophoresis and Hybridization

Restriction enzymes were obtained from Takara Shuzo Co., Ltd. (Kyoto, Japan) except for Msp I, which was purchased from New England Biolaboratory. After cleavage with restriction enzymes, DNA samples were electrophoresed on 0.7% horizontal agarose gels and transferred to a nitrocellulose sheet, as described by Southern (1975).

Total cellular RNA was denatured, as described (McMaster and Carmichael, 1977), electrophoresed on 1.1% agarose gels and transferred to DBM paper accordingly (Alwine et al., 1977). For the dot blot hybridization, nondenatured RNA was spotted directly onto nitrocellulose filter (Thomas, 1980).

The recombinant plasmid which carries the 8.2 kb M-MuLV genome at the Hind III site and which was cloned by J. W. Hoffmann, was a generous gift from R. A. Weinberg. The DNA fragment of 8.2 kb containing the entire sequence of M-MuLV was purified by agarose gel electrophoresis. M-MuLV DNA was further digested by Bam HI and Sma I. The Hind III–Bam HI fragment of 1.65 kb was recovered from agarose gel after electrophoresis. The Sma I fragment of 316 bases was electrophoresed through a polyacrylamide gel. Recombinant plasmids carrying 18S and 28S rRNA gene of the mouse were a kind gift from R. Kominami (Urano et al., 1980; Kominami et al., 1982). These DNAs were labeled by the nick translation procedure (Maniatis et al., 1975). Specific activities of the probes were 1–4 × 10^8 cpm/μg.

Acknowledgments

We thank Drs. R. A. Weinberg and J. W. Hoffmann for the generous gift of recombinant plasmid, M. Yoshida for pertinent advice on biochemical techniques, A. Ishimoto for providing M-MuLV stock.

Prof. S. Numa, in whose laboratory part of this investigation was carried out, R. Kominami for provision of plasmids carrying rRNA genes, K. Sekikawa from whom we obtained the F-9 cells, M. Ohara for reading the manuscript and Y. Fukuda for secretarial services. This work was supported by a grant-in-aid from the Ministry of Education, Science, and Culture, Japan.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 14, 1982; revised January 25, 1983

References

Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Nat. Acad. Sci. USA *74*, 5350–5354.

Artzt, K. and Jacob, F. (1974). Absence of serologically detectable H-2 on primitive teratocarcinoma cells in culture. Transplantation *17*, 632–634.

Bernstine, E. G., Hooper, M. L., Granchamp, S. and Ephrussi, B. (1973). Alkaline phosphatase activity in mouse teratoma. Proc. Nat. Acad. Sci. USA *70*, 3899–3903.

Chan, H. W., Bryan, T., Moore, J. L., Staal, S. P., Rowe, W. P. and Martin, M. A. (1980). Identification of ecotropic proviral sequences in inbred mouse strains with a cloned subgenomic DNA fragment. Proc. Nat. Acad. Sci. USA *77*, 5779–5783.

Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. and Rowe, W. P. (1974). Qualitative and quantitative studies on AKR-type murine leukemia virus sequences in mouse DNA. Cold Spring Harbor Symp. Quant. Biol. *39*, 1085–1101.

Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry *18*, 5294–5299.

Dolberg, D. S., Bachelet, L. T. and Fan, H. (1981). Endogenous type C retroviral sequences of mice are organized in a small number of virus-like classes and have been acquired recently. J. Virol. *40*, 96–106.

Epstein, C. J., Smith, S., Travis, B. and Tucker, G. (1978). Both X chromosomes function before visible X-chromosome inactivation in female mouse embryos. Nature *274*, 500–503.

Friedrich, T. D. and Lehman, J. M. (1981). The state of simian virus 40 DNA in the embryonal carcinoma cells of the murine teratocarcinoma. Virology *110*, 159–166.

Gautsch, J. W. (1980). Embryonal carcinoma stem cells lack a function required for virus replication. Nature *285*, 110–112.

Gautsch, J. W. and Wilson, M. (1983). Delayed *de novo* methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. Nature, in press.

Goeddel, D. V., Yansura, D. G. and Caruthers, M. H. (1977). Binding of synthetic lactose operator DNAs to lactose repressors. Proc. Nat. Acad. Sci. USA *74*, 3292–3296.

Groudine, M., Eisenman, R. and Weintraub, H. (1981). Chromatin structure of endogenous retroviral genes and activation by an inhibitor of DNA methylation. Nature *292*, 311–317.

Hartley, J. W. and Rowe, W. P. (1975). Clonal cell line from a feral mouse embryo which lacks host range restriction for murine leukemia virus. Virology *65*, 128–134.

Hirt, B. (1967). Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. *26*, 365–369.

Hoffman, J. W., Steffen, D., Gusella, J., Tabin, C., Bird, S., Cowing, D. and Weinberg, R. A. (1982). DNA methylation affecting the expression of murine leukemia proviruses. J. Virol. *44*, 144–157.

Heubner, K., Tsuchida, N., Green, C. and Croce, C. M. (1979). A murine teratocarcinoma stem cell line carries suppressed oncogenic virus genomes. J. Exp. Med. *150*, 392–405.

Ihle, J. N., Joseph, D. R. and Domotor, J. J. (1979). Genetic linkage of C3H/HeJ and BALB/c endogenous ecotropic C-type viruses to phosphoglucomutase-1 on chromosome 5. Science *204*, 71–73.

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proc. Nat. Acad. Sci. USA *73*, 1260–1264.

Jaenisch, R., Fan, H. and Croker, B. (1975). Infection of preimplantation mouse embryos and of new born mice with leukemia virus: tissue distribution of viral DNA and RNA leukemogenesis in the adult animal. Proc. Natl. Acad. Sci. USA *72*, 4008–4012.

Jakob, H., Boon, T., Gillard, J., Nicholas, J. F. and Jacob, F. (1973). Teratocarcinoma de la souris: isolement, culture et proprietes de cellules a potentialites multiples. Ann. Microbiol. (Paris). *124B*, 269–282.

Kleinsmith, L. J. and Pierce, G. B., Jr. (1964). Multipotentiality of single embryonal carcinoma cells. Cancer Res. *24*, 1544–1551.

Kominami, R., Mishima, Y., Urano, Y., Sakai, M. and Muramatsu, S. (1982). Cloning and determination of the transcription termination site of ribosomal RNA gene of the mouse. Nucl. Acids Res. *10*, 1963–1979.

Kozak, C. A. and Rowe, W. P. (1979). Genetic mapping of the ecotropic murine leukemia virus-inducing locus of BALB/c mouse to chromosome 5. Science *204*, 69–71.

Kozak, C. A. and Rowe, W. P. (1980). Genetic mapping of the ecotropic virus-inducing locus *Akv-2* of the AKR mouse. J. Exp. Med. *152*, 1419–1423.

Lehman, J. M., Speers, W. C., Swartzendruber, D. E. and Pierce, G. B. (1974). Neoplastic differentiation: characteristics of cell lines derived from a murine teratocarcinoma. J. Cell. Physiol. *84*, 13–28.

Lin, S. Y. and Riggs, A. (1972). *Lac* operator analogues: bromodeoxyuridine substitution in the *lac* operator affects the rate of dissociation of the *lac* repressor. Proc. Nat. Acad. Sci. USA *69*, 2574–2576.

Lin, S. Y. and Riggs, A. (1976). The binding of *lac* repressor and the catabolite gene activator protein to halogen-substituted analogues of poly (dA-T). Biochim. Biophys. Acta. *432*, 185–191.

Maniatis, T., Jeffrey, A. and Kleid, D. G. (1975). Nucleotide sequence of the rightward operator of phage. Proc. Nat. Acad. Sci. USA *72*, 1184–1188.

Martin, G. R., Epstein, C. J., Travis, B., Tucker, G., Yatziv, S., Martin, D. M., Clift, S. and Cohen, S. (1978) X-chromosome inactivation during differentiation of female teratocarcinoma stem cells in vitro. Nature *271*, 329–333.

McBurney, M. W. and Strutt, B. J. (1980). Genetic activity of X chromosomes in pluripotent female teratocarcinoma cells and their differentiated progeny. Cell *21*, 357–364.

McMaster, G. K. and Carmichael, G. G. (1977). Analysis of single- and double-stranded nucleic acids on polyacrylamide and agrose gels by using glyoxal and acridine orange. Proc. Nat. Acad. Sci. USA *74*, 4835–4838.

Miller, O. J., Schnedl, W., Allen, J. and Erlenger, B. F. (1974). 5-Methylcytosine localised in mammalian constitutive heterochromatin. Nature *251*, 636–637.

Mintz, B. and Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proc. Nat. Acad. Sci. USA *72*, 3585–3589.

Niwa, O. and Sugahara, T. (1981). 5-Azacytidine induction of mouse endogenous type C virus and suppression of DNA methylation. Proc. Nat. Acad. Sci. USA *78*, 6290–6294.

Niwa, O., Decleve, A., Lieberman, M. and Kaplan, H. S. (1973). Adaptation of plaque assay methods to the in vitro quantitation of the radiation leukemia virus. J. Virol. *12*, 68–73.

Peries, J., Alves-Cardoso, E., Canivet, M., Devons-Guillemin, M. C. and Lasneret, J. (1977). Lack of multiplication of ecotropic murine C-type viruses in mouse teratocarcinoma primitive cells. J. Nat. Cancer Inst. *59*, 463–465.

Rezin, A. and Ceder, H. (1977). Distribution of 5-methylcytosine in chromatin. Proc. Nat. Acad. Sci. USA *74*, 2725–2728.

Rowe, W. P., Pugh, W. E. and Hartley, J. W. (1970). Plaque assay techniques for murine leukemia viruses. Virology *42*, 1136–1139.

Rowe, W. P., Hartley, J. W. and Besmer, T. (1972). Genetic mapping of a murine leukemia virus–inducing locus of AKR mice. Science *178*, 860–862.

Segal, S. and Khoury, G. (1979). Differentiation as a requirement for simian virus 40 gene expression in F-9 embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *76*, 5611–5615.

Sekikawa, K. and Levine, J. (1981). Isolation and characterization of polyoma host range mutants that replicate in nullipotential embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *78*, 1110–1104.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. *98*, 503–518.

Speers, W. C., Gautsch, J. W. and Dixon, F. J. (1960). Silent infection of murine embryonal carcinome cells by Moloney murine leukemia virus. Virology *105*, 241–244.

Stewart, C. L., Stuhlman, H., Jähner, D. and Jaenisch, R. (1982) *De novo* methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Nat. Acad. Sci. USA *79*, 4098–4102.

Strickland, S. and Mahdavi, V. (1978). The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell *15*, 393–403.

Stuhlmann, H., Jähner, D. and Jaenisch, R. (1981). Infectivity and methylation of retrovirus genomes is correlated with expression in the animal. Cell *26*, 221–232.

Swartzendruber, D. E. and Lehman, J. M. (1975). Neoplastic differentiation: interaction of simian virus 40 and polyoma virus with murine teratocarcinoma cells in vitro. J. Cell Physiol. *85*, 179–188.

Swartzendruber, D. E., Friedrich, T. D. and Lehman, J. M. (1977). Resistance of teratocarcinoma stem cells to infection with simian virus 40: early events. J. Cell Physiol. *93*, 25–30.

Teich, N. M., Weiss, R. A., Martin G. R. and Lowy, D. R. (1977). Virus infection of murine teratocarcinoma stem cell lines. Cell *12*, 973–982.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Nat. Acad. Sci. USA *77*, 5201–5205.

Urano, Y., Kominami, R. and Muramatsu, S. (1980). The nucleotide sequence of the putative transcription initiation site of a cloned ribosomal RNA gene of the mouse. Nucl. Acids Res. *8*, 6043–6058.

van der Putten, H., Quint, W., Verma, I. M. and Berns, A. (1982). Moloney murine leukemia virus–induced tumors: recombinant proviruses in active chromatin regions. Nucl. Acids Res. *10*, 577–592.

TAB 6

This material may be protected by Copyright law (Title 17 U.S. Code)

ELSEVIER

Gene 261 (2000) 181–188

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Retroviruses in foreign species and the problem of provirus silencing☆

J. Svoboda[a,*], J. Hejnar[a], J. Geryk[a], D. Elleder[a], Z. Vernerová[b]

[a]*Institute of Molecular Genetics, Academy of Sciences of the Czech Republic, Flemingovo n. 2, 166 37 Prague 6, Czech Republic*
[b]*Department of Pathology, First Faculty of Medicine, Charles University, Prague, Czech Republic*

Received 29 May 2000; received in revised form 18 August 2000; accepted 5 October 2000
Received by G. Bernardi

Abstract

Retroviruses are known to integrate in the host cell genome as proviruses, and therefore they are prone to cell-mediated control at the transcriptional and posttranscriptional levels. This plays an important role especially after retrovirus heterotransmission to foreign species, but also to differentiated cells. In addition to host cell-mediated blocks in provirus expression, also so far undefined host specificities, deciding upon the pathogenic manifestation of retrovirus heterotransmission, are in play. In this respect, we discuss especially the occurrence of wasting disease and immunodeficiency syndrome, which we established also in avian species using avian leukosis virus subgroup C (ALV-C) inoculated in mid-embryogenesis in duck or chicken embryos. The problem of provirus downregulation in foreign species or in differentiated cells has been in the recent years approached experimentally. From a series of observations it became apparent that provirus downregulation is mediated by its methylation, especially in the region of proviral enhancer-promoter located in long terminal repeats (LTR). Several strategies have been devised in order to protect the provirus from methylation using LTR modification and/or introducing in the LTR sequence motifs acting as antimethylation tags. In such a way the expression of retroviruses and vectors in foreign species, as well as in differentiated cells, has been significantly improved. The complexity of the mechanisms involved in provirus downregulation and further possibilities to modulate it are discussed. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Retrovirus heterotransmission; Retroviral vector; Methylation; CpG; Long terminal repeat(s); Histone deacetylase

1. Introduction

The topic of this review is focused on selected aspects of retrovirus heterotransmission, but also touches the problem of retroviral genome silencing in differentiated and other types of cells. In case of heterotransmission, many host cell factors are in play, which decide whether or to what degree the retroviral genome will be expressed and what pathogenic consequences may be triggered. As is known generally, the retroviral genome becomes integrated in the cell genome as a provirus and, therefore, it is not surprising that it is highly influenced by the host-cell gene-regulation machinery. In discussing such downregulation of the provirus we are in fact dealing with post-integration blocks in provirus expression, the nature of which is epigenetic and mediated by new host cell factors. The unusual cell milieu, in concert with which the virus has not been evolving, can be lacking some factors like those enabling viral RNA export from the nucleus, or provide unusual factors like those changing the viral RNA splicing. However, of main importance is cell transcriptional regulation, which in many cases leads to provirus silencing. It is, therefore, not surprising that in phylogenetically distant host cells the provirus can integrate, but in many cases does not produce an infectious progeny. We call such host cells non-permissive, in contrast to permissive cells where formation of infectious virions takes place.

Non-permissiveness to retroviral infection has been for a

Abbreviations: ALV, avian leukosis virus(es); ALV-B, ALV-C, ALV-D, avian leukosis virus subgroup B, C, and D, respectively; *aprt*, adenosine-phosphoribosyltransferase gene; β-*geo*, fused β-galactosidase and neomycin resistance gene; BLV, bovine leukemia virus; CAT, chloramphenicol acetyltransferase; cHS4, chicken hypersensitive site 4; EC, embryonic carcinoma; GFP, green fluorescence protein; HDAC, histone deacetylase; HIV, human immunodeficiency virus; IFN-SAR, interferon scaffold attachment region; LCR, locus control region(s); LTR, long terminal repeat(s); MeCP, methyl-CpG-binding protein; MEL, murine erythroleukemia; MLV, murine leukemia virus(es); MSV, murine sarcoma virus; NCR, negative control region; *neo*, neomycin resistance gene; PR RSV, Prague strain of RSV; RSV, Rous sarcoma virus; SIV, simian immunodeficiency virus; SIVcpz, chimpanzee SIV; SIVsm, sooty mangabey SIV; TSA, trichostatin; X-MLV, xenotropic MLV

☆ Extended version of the lecture "Modification of retrovirus pathogenicity by transspecies transfer" presented at the workshop "Microbial Variation and Evolution" at Ischia, Italy (organizers G. Bernardi and B. Fantini).

* Corresponding author. Tel.: +42-2-243-10-238; fax: +42-2-243-10-955.
E-mail address: svoboda@img.cas.cz (J. Svoboda).

0378-1119/00/$ - see front matter © 2000 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(00)00481-9

long time ascribed to foreign species host cells. Interestingly enough, non-permissiveness to retroviral infection marks also some cells of the virus species origin. This has been recognized especially in cases of cultured differentiated cells. In such a way, surmounting hurdles negatively influencing provirus expression became of importance for an efficient application of retroviral vectors for gene therapy of developmentally committed stem cells and differentiated cells. Therefore, we also discuss the provirus fate in some defined cases of differentiated cells that originate from the same species as the retrovirus. There are common features shared among all these situations, pointing to a role of cell-mediated gene silencing as an important factor that superimposes upon provirus expression.

2. Trans-class retrovirus heterotransmission

This extreme situation was first achieved in the case of chicken Rous sarcoma virus (RSV) strains transmitted first to rodents and later also to other mammalian species, including monkeys (rev. Svoboda, 1986). RSV genome expression in mammalian cells is governed by a series of factors, especially by flanking DNA sequences, their richness in GC (Fincham and Wyke, 1991; Rynditch et al., 1991), but also posttranscriptional steps are involved, which are blocked in mammalian cells (rev. Svoboda, 1998). More recent progress indicates that of importance are RSV LTR, which are prone to methylation in mammalian cells.

It was shown experimentally that a reporter gene driven by in vitro methylated RSV LTR is more efficiently suppressed in mammalian cells as compared to chicken cells (Hejnar et al., 1999). In spite of the fact that there are 16 CpGs in the Prague strain of Rous sarcoma virus (PR RSV) LTR and multiple CpGs are present in leader sequences, methylation of one CCGG *Hpa*II site, located downstream of the promoter region but close to the single provirus transcriptional start, was sufficient to produce dramatic reporter downregulation in mammalian cell lines. It should be noted that these experiments were done using transient transfection of methylated proviral DNAs and unmethylated controls. Expression of unmethylated LTR was comparable in both avian and non-permissive mammalian cells, suggesting that both types of cells harbor sufficient transcriptional machinery required by RSV LTR. Silencing of RSV proviruses is therefore a post-integration event.

The significance of LTR methylation for provirus downregulation has been recently approached using two strategies. The first one, schematically shown in Fig. 1, implies insertion of four canonical Sp1 binding sites in the RSV LTR enhancer region using *Eco*RI sites. Such reconstruction has been documented and discussed in detail (Machoň et al., 1998), and it was found that Sp1 insertion significantly increases LTR-driven chloramphenicol acetyltransferase (CAT) reporter gene expression, especially in hamster cells using both transient and stable transfection assays. This is in agreement with findings revealing that Sp1 binding sites represent the critical part of sequences acting as antimethylation tags, as was documented clearly in the case of a CpG island containing 1.7 kb DNA located in front of the adenosine-phosphoribosyltransferase (*aprt*) gene (Mummaneni et al., 1993; Brandeis et al., 1994; Macleod et al., 1994). Acquisition of an Sp1 binding site by mutation in the LTR enhancer region was shown previously to activate MLV (murine leukemia virus) transcription in embryonic carcinoma (EC) cells (Prince and

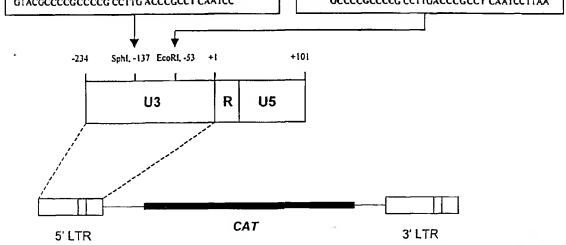


Fig. 1. Scheme of the CAT reporter vector employed for insertion of Sp1 sites into LTR sequence and transfection experiments. Sites of insertion (*Sph*I and *Eco*RI) are denoted by arrows. Sp1 binding sites within the inserted sequences are in bold.

Rigby, 1991). In a similar way, transcriptional activation of human endogenous retrovirus (HERV-H) was accomplished by Sp1 occurrence in either LTR enhancer (Nelson et al., 1996) or promoter (Sjøttem et al., 1996; Anderssen et al., 1997). The second approach is based on the incorporation of the mouse *aprt* gene CpG island immediately upstream to the RSV LTR-driven fused β-galactosidase and neomycin resistance (β-*geo*) proviral reporter (see Fig. 2). After transfection of hamster cells and selection for neomycin resistance, such construction ensured stable transcriptional activity in a reasonable number of cell clones, whereas unprotected proviral reporters are inactivated. This effect can be attributed to the antimethylation protection of the CpG island for two reasons; first, the active proviruses remained unmethylated within the 5′ LTR as evidenced by the bisulphite sequencing technique, second, the transcription efficiency of the RSV LTR is not increased in the presence of the CpG island.

Thus, both approaches led to the conclusion that proviral LTR could be protected from DNA methylation in the foreign species host. Optimization of these protective strategies might open the way to construct improved RSV-based vectors for gene transfer, more suitable for expression in mammalian cells and without the risk of infectious retroviral progeny. Retroviral vector producer cells (helper cells) are an important object for such protection as well. These cells have been shown genetically unstable due to the methylation of integrated helper proviral constructions. Designing a helper virus to overcome cellular DNA methylation may therefore improve vector production (Young et al., 2000).

There exists also another way of trans-species retrovirus transmission utilizing xenotropic murine leukemia viruses (X-MLV) (rev. Levy, 1978). It is interesting that these viruses can replicate in some avian cells such as duck cells, but not in others like chicken cells (Levy, 1977). X-MLV can provide envelope components to pseudotype RSV virions and such pseudotypes transform and replicate in duck but not in chicken cells. In vivo inoculation of X-MLV or murine sarcoma virus (MSV) pseudotyped by X-MLV was performed in duck embryos or newborn ducklings (Levy et al., 1982). Evidence of virus persistence has been obtained, but not convincing data concerning their pathogenesis. In mammalian cells, X-MLV-pseudotyped RSV can also replicate in the presence of X-MLV first to a low titer, which increases with passages. Furthermore, an envelope component of ALV-C phenotypically mixed with X-MLV was detected after passaging in mammalian cells.

It is not known how X-MLV contributes to RSV replication in mammalian cells. Obviously, it provides at least a part of the Env glycoprotein required for penetration to mammalian cells. However, additional factors complementing the non-permissiveness of mammalian cells to RSV should be in play. Due to its ability to multiply in such cells, X-MLV can increase and facilitate some posttranscriptional steps such as viral RNA proper splicing and export from the nucleus and/or further steps involving cleavage of viral protein precursors and virion assembly. These facilitating effects of X-MLV should be synchronized with RSV infection, because superinfection of an already RSV-transformed mammalian cell, containing functional proviruses rescuable by fusion with chicken fibroblasts, does not result in virus production (Levy, 1977).

Because X-MLV contribution to RSV replication in mammalian cells increases with passages, a possibility of genetic exchange between these viruses should be taken into account. In every case, these questions should be reinvestigated using presently available efficient tools of molecular biology.

3. Retrovirus heterotransmission among species within the same class

One of the first thoroughly analyzed successful heterotransmissions was achieved by Duran-Reynals (1942), who showed that RSV produces both early and late appearing tumors when inoculated in young ducks. Since then, many

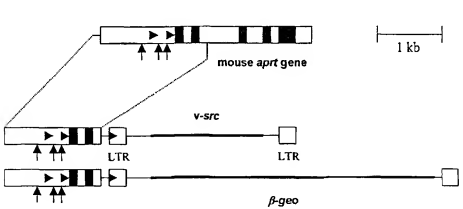


Fig. 2. Cloning of the *aprt* gene CpG island and RSV-based reporter proviruses. Filled boxes represent exons of the *aprt* gene. Vertical arrows denote the position of three Sp1 sites in the CpG island. Filled arrowheads denote transcription starts of the *aprt* gene and 5′ LTR.

Table 1
Trans-species retrovirus transmission

Virus	Species of origin	Original pathogenicity	Transmitted to species	New species pathogenicity	References
SIV_{cpz}	Chimpanzees	Zero to very low	Human	AIDS (HIV-1)	Gao et al., 1999; Weiss and Wrangham, 1999.
SIV_{sm}	Sooty mangabey	Zero to very low	Human	AIDS (HIV-2)	Hirsch et al., 1989; Gao et al., 1992; Sharp et al., 1995; Chen et al., 1996.
BLV	Cattle	Lymphocytosis, leukosis, B-tropic	Rabbit	Wasting disease, immunodeficiency	Burny et al., 1985; Altanerova et al., 1989; Wyatt et al., 1989; Kucerova et al., 1999.
Friend MLV	Mice	Erythroleukemia	Rat	Bone marrow suppression, $Thy1^+$ cell reduction	Mazgareanu et al., 1998.
ALV-C	Chicken	Anemia, probably immunodeficiency	Duck	Wasting disease, immunodeficiency, anemia	Karakoz et al., 1980;
ALV-B ALV-D		Anemia			Smith and Schmidt, 1982.

other retroviruses, including these of mammalian origin, were experimentally transmitted among different mammalian species. These transmissions were monitored mainly by virus oncogenic activity and virus persistence. Because of generally low retrovirus replication in foreign species, additional pathogenic virus activity, such as immunosuppression, usually do not appear. In many cases heterotransmission produced the same symptoms as in the species of the virus origin, but there are well-documented cases of changed virus pathogenicity. Further, we shall deal mainly with the symptoms of wasting disease accompanied by immunodeficiency resulting in increased susceptibility to various infectious agents.

The simian immunodeficiency viruses (SIV) highlight this situation. Generally, in the monkey species, in which SIV is indigenous, it replicates efficiently but does not produce any pathogenic changes. However, when transmitted to some other monkey species, SIV produces the immunodeficiency syndrome. Of special interest is SIV heterotransmission to humans. As is summarized in Table 1, good evidence based on molecular biology and epidemiology has been provided documenting that both human immunodeficiency virus 1 (HIV-1) and HIV-2 represent a consequence of respective SIV_{sm} (sooty mangabey SIV) and SIV_{cpz} (chimpanzee SIV) heterotransmission to humans. There is no doubt about the significance of this finding, which should be taken as a warning against potential danger of retrovirus transgression of species barriers associated with fulminating pathogenic changes. In order to understand these events, comparative data obtained with other members of the retrovirus family should be evaluated.

Interesting observations were made in the case of bovine leukemia virus (BLV) (Table 1). This virus responsible for cattle leukosis when transferred to newborn rabbits triggers clear symptoms of immunodeficiency. Because such a response was not found in other infected species, these observations indicate that rabbits respond to BLV inherently, in an unusual way.

In murine leukemia viruses, variants capable to produce immunodeficiency preferably have been isolated. This includes both Moloney MLV (Saha et al., 1994) and Friend leukemia complex (Faxvaag et al., 1993). In addition, as given in Table 1, Friend MLV transmission to newborn rats resulted in altered pathogenicity characterized by suppression of bone marrow cells, manifesting itself as reduced numbers of $Thy1^+$ cells.

Avian leukosis viruses have not been thoroughly investigated from the point of view of their immunopathogenicity in foreign avian species. As given in Table 1, ALV subgroup C were studied using intraembryonic inoculation both in chicken and ducks. According to data obtained so far, this subgroup produces symptoms of anemia in both species. However, in heterologous duck hosts, a fatal wasting disease together with conspicuous atrophy of the thymus tissue starting the first week after hatching was observed both by a decrease in the relative thymus to body weight, histologically characterized by clear depletion of the thymus cortical layer (Fig. 3) (Stepanets et al., 2000). Microscopically, bursa Fabricii, which constitutes a special B-cell producing organ in birds, also displayed the cortical layer depletion. In agreement with this observation, production of humoral antibodies against *Brucella abortus* antigens was significantly decreased in infected animals, which confirms that immunodeficiency is involved. The nature of this immunodeficiency is being investigated and the character of specific T- and other cell alteration in thymus is monitored by specific antibodies. In addition, changes in lymphoid organs in young chickens intraembryonally inoculated with different ALV subgroups should be investigated. According to our preliminary data, ALV-C does produce symptoms of thymus involution even in chickens. Thus ALV-C provides a suitable comparative system for establishing ways leading to the immunodeficiency syndrome both in homologous and heterologous hosts.

There is not a simple answer to the question why retroviruses in some species behave pathogenically or produce

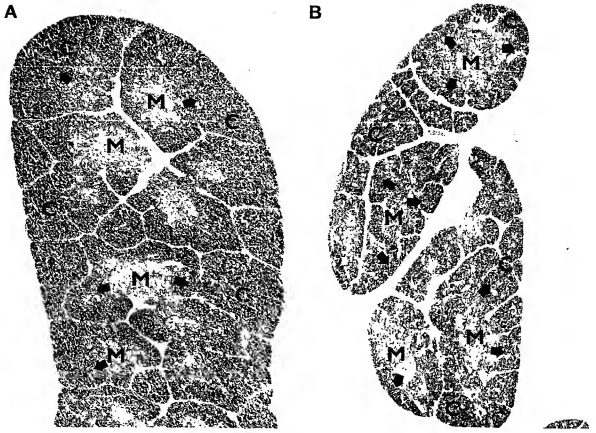


Fig. 3. Thymus sections of 7-day-old ducklings. (A) Thymus of a control animal, mock-injected with tissue culture medium. (B) Thymus of an animal infected in mid-embryogenesis with 10^5 ALV-C infectious virus. Thymic cortical layer is visibly depleted. C, cortex; M, medulla; arrowheads mark the interim between cortex and medulla. Stained with hematoxylin-eosin. Magnification 41×.

new or accentuated pathogenic symptoms. In some cases, such as Friend MLV, selection of virus mutants might be involved, but this seems not to be a general situation, because altered pathogenicity appears very soon after virus inoculation and is not correlated with certain specific virus gene alteration. Therefore other, especially host-specific factors should be taken into account. It is not known so far which of them play a decisive role. Retrovirus toxicity is in the first step related to the outcome of the viral envelope (product of the *env* gene) interaction with target cells, immune and bone-marrow cells included. Therefore, if an infected cell harbors unusually expressed receptors and co-receptors as well as factors required for further steps of virus penetration, they may undergo deterioration or apoptosis. As was discussed in relation to HIV (Fauci, 1996), the degree of cytokine activity and the outcome of interrelation among different cytokine pathways contribute to virus expression and its pathogenicity.

In spite of the fact that heterotransmission of SIV took place probably several times, we are lacking the exact information about this process. In contrast to the experimental systems, natural conditions are probably different, because only low amounts of virus or better virus-infected cells from a donor might have mediated heterotransmission. Such a situation has not been modeled in detail, but it can be predicted that under such conditions the virus becomes efficiently expressed only rarely. Therefore, there might be a bridge connecting trans-class and trans-species retrovirus heterotransmission.

Some experimental data obtained from long-term follow-up of retrovirus persistence (Trejbalová et al., 1999) suggest that provirus silencing discussed in the first and also next chapter might play a role in the control of low-dose infection and long-term retrovirus persistence.

4. Provirus silencing and expression in differentiated cells

In order to investigate the role of cell differentiation in relation to retrovirus infection, several models have been designed, but most thoroughly were analyzed EC cells, which differentiate in vitro. The entrée to this problem was given by the group of Jaenisch (Stewart et al., 1982).

who discovered that MLV integrated in EC cells becomes methylated and unexpressed. In a series of papers from different laboratories (reviewed by Challita et al., 1995) it was established that in order to ensure MLV expression in EC, introduction of a transcription factor Sp1 binding site into LTR is required together with inactivation of the negative control region (NCR) in LTR as well as of the region of the primer binding site. As the third negatively acting element, one out of two direct repeats in LTR was recognized (Hawley et al., 1994). There is still room left for MLV improved expression, as exemplified by insertion into LTR of an antimethylation fragment from the region upstream of the Thy1 gene (Challita et al., 1995).

All the above-mentioned modifications act synergistically. Recently, MLV LTR lacking known negative elements due to deletion spanning most of the 5′ end LTR portion has been constructed (Osborne et al., 1999). As a result of such a deletion, about half of 13 CpG sites within LTR were also lost. Such truncated LTR, essentially stripped of enhancer elements, was employed for generation of a retroviral vector harboring the neomycin resistance *(neo)* reporter gene equipped with internal β-globin promoter, which after infection of EC cells displayed expression in 70% of cells, the highest efficiency obtained so far.

We have not yet reached the end of the journey to optimal retroviral vector function in differentiated cells. It is still possible that additional, so far undefined sequences should be inactivated, altered or inserted. Of special importance might be CpG dinucleotides present in LTR, especially at the start of transcription. Elements acting as antimethylation signals, interfering with silencers or ensuring position-independent gene expression, such as locus control regions (LCR), should be tested for their ability to ensure retrovirus or retroviral vector expression in differentiated cells. Recently, the chicken hypersensitive site 4 (cHS4) of the chicken globin LCR, acting as an insulator, when cloned into MLV LTR was shown to increase the probability of integrated proviruses expression and to decrease the level of de novo methylation of the 5′LTR in murine erythroleukemia (MEL) cells (Rivella et al., 2000). In addition, the human β interferon scaffold attachment region (IFN-SAR), when inserted in retroviral LTR, prevented its methylation and ensured vector expression in a stably transfected line of human T cells. The vector expression has been kept for several months and included also multiple proviral copies (Agarwal et al., 1998; Dang et al., 2000). We can therefore stress the point that the problem of permissiveness of differentiated cells to retrovirus infection goes far beyond EC cells and that other differentiated cells, such as hematopoietic or hepatic cells, should probably require not only prevention of vector downregulation, but even some more specific cell changes like activation of steps triggering the cell cycle (rev. Emerman, 2000).

Dealing with provirus silencing in differentiated cells we focused on provirus methylation as an epigenetic DNA modification described repeatedly in conjunction with provirus downregulation. However, it is not clear so far whether provirus methylation acts as the primary cause or whether it only conserves transcriptional repression. In provirus methylation studies, usually a general increase in CpG methylation has been measured. However, some more precise data have been obtained showing that methylation of one particular CpG has a decisive effect. This is the case of HIV, where methylation of only one CpG at position −143 (in the vicinity of NF-κB and Sp1 binding sites) in HIV LTR results in 70% inhibition of the reporter expression (Bednarik et al., 1990; Schulze-Forster et al., 1990). Similarly, single CpG site methylation in the RSV LTR U5 region nearby the transcription start leads to a clear decrease in provirus expression (Hejnar et al., 1999). As discussed later, the density of CpGs is also of importance.

It has been recognized on other gene models that methylation provides a signal for association with methyl-CpG-binding protein 2 (MeCP2), which, through the adaptor protein Sin3A, recruits the histone deacetylase (HDAC) (rev. Razin, 1998; Ng and Bird, 1999; Knoepfler and Eisenman, 1999). A similar situation was disclosed in the case of MeCP1, which produces a complex composed of the MBD protein containing the methyl-CpG-binding domain and two members of the HDAC family (rev. Bird and Wolffe, 1999). Thus, there is a proven link between CpG methylation and chromatin deacetylation.

How are these findings related to provirus silencing? It should be noted that MeCP1 has been already shown to bind to methylated LTR of myeloproliferative sarcoma virus, suppressing its LTR activity (Boyes and Bird, 1991). This problem has been recently approached by Lorincz et al. (2000). They employed the MLV LTR-driven green fluorescence protein (GFP) gene and followed GFP expression after infection of MEL cells with this vector. Cell clones that displayed silenced GFP were isolated and it was disclosed that early after silencing the proviruses became methylated to a low degree. Such clones could have been reactivated by trichostatin (TSA), which inhibits HDAC-MeCP2 complexes. During prolonged cell cultivation the vector has been increasingly methylated and in the hypermethylated state provirus expression was induced only by combination of TSA and 5-azacytidine (5-azaC), which acts as a demethylation agent. These results suggest that provirus downregulation is a dynamic process and that the possibility of its reactivation depends upon the density of methylation.

All the data obtained so far point to the important role of methylation in provirus silencing in general, and therefore strategies preventing methylation of retroviral vectors as well as putative blocking of methyl-CpG-binging proteins should contribute to more efficient gene therapy applied to differentiating or differentiated cells.

There is, in addition, an important question: whether and why retroviruses are more efficiently recognized and silenced than any foreign DNA introduced in a genome. In other words, does there exist a cell genome surveillance mechanism (analogous but not homologous to immunity)

that ensures downregulation of retroviral sequences? Methylation has been already proposed to fulfil such a duty (Doerfler, 1991; Yoder et al., 1997). However, some signals should be involved that attract methyltransferase to an integrated provirus, or even to a specific DNA structure common to viral integration intermediates (Bestor, 1987). Such signals might be provided by flanking chromosomal sequences or by the proviral structure itself. Especially LTR could be recognized as unusual direct repeats. Similar structures in lower eukaryotes trigger gene silencing (rev. Wolffe and Matzke, 1999). Furthermore, MLV LTR binding zinc finger transcription factor YY-1 (Flanagan et al., 1992) was shown to represent a homolog of Sin3 (Yang et al., 1996), which has been already identified as a protein complexing with HDAC involved in formation of transcriptionally inactive chromatine. It should be also taken into account that in some heterologous and differentiated cells, proteins activating LTR might be underrepresented or that such cells produce altered isomorphic proteins, which could be inactive and/or could interfere with factors required for LTR activation. Therefore, retroviral genomes should be screened also from the point of view of sequences and DNA-protein complexes which might contribute to provirus silencing, of course, in context with flanking chromosomal DNA. New techniques such as inverse polymerase chain reaction should facilitate this demanding task.

Acknowledgements

The authors highly acknowledge the excellent editorial help of Šárka Takáčová. These studies, which involve several interconnected areas of research, were supported by Grant No. 1R03 TW00 155-01 A1 (Fogarty International Research Collaboration Award), Grants No. 301/94/0713, No. 203/96/0883, No. 312/96/K205, No. 312/98/0825 and No. 312/97/P082 from the Grant Agency of the Czech Republic.

References

Agarwal, M., Austin, T.W., Morel, F., Chen, J., Böhnlein, E., Plavec, I., 1998. Scaffold attachment region-mediated enhancement of retroviral vector expression in primary T cells. J. Virol. 72, 3720–3728.

Altunerova, V., Ban, J., Altaner, C., 1989. Induction of immune deficiency syndrome in rabbits by bovine leukaemia virus. AIDS 3, 755–758.

Andersen, S., Sjettem, E., Svineng, G., Johansen, T., 1997. Comparative analysis of LTR of the ERV-H family of primate-specific retrovirus-like elements isolated from mammoset African green monkey, and man. Virology 234, 14–30.

Bednarik, D.P., Cook, J.A., Pitha, P.M., 1990. Inactivation of the HIV LTR by DNA CpG methylation Evidence for a role in latency. EMBO J. 9, 1157–1164.

Bestor, T.H., 1987. Supercoiling-dependent sequence specificity of mammalian DNA methyltransferase. Nucleic Acids Res. 15, 3835–3843.

Bird, A.P., Wolffe, A.P., 1999. Methylation-induced repression - belts, braces, and chromatin. Cell 99, 451–454.

Boyes, J., Bird, A., 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. Cell 64, 1123–1134.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., Cedar, H., 1994. Sp1 elements protect a CpG island from de novo methylation. Nature 371, 435–438.

Burny, A., Bruck, C., Cleuter, Y., Couez, D., Deschamps, J., Ghysdael, J., Gregoire, D., Kettmann, R., Mammerickx, M., Marbaix, G., 1985. Bovine leukaemia virus, a versatile agent with various pathogenic effects in various animal species. Cancer Res. 45, 4578s–4582s.

Challita, P.-M., Skelton, D., El-Khoueiry, A., Yu, X.-J., Weinberg, K., Kohn, D.B., 1995. Multiple modifications in *cis* elements of the long terminal repeat of retroviral vectors lead to increased expression and decreased DNA methylation in embryonic carcinoma cells. J. Virol. 69, 748–755.

Chen, Z., Telfer, P., Gettie, A., Reed, P., Zhang, L., Ho, D.D., Marx, P.A., 1996. Genetic characterization of new West African simian immunodeficiency virus SIVsm: geographic clustering of household-derived SIV strains with human immunodeficiency virus type 2 subtypes and genetically diverse viruses from a single feral sooty mangabey troop. J. Virol. 70, 3617–3627.

Ding, Q., Auten, J., Plavec, I., 2000. Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector. J. Virol. 74, 2671–2678.

Doerfler, W., 1991. Patterns of DNA methylation - evolutionary vestiges of foreign DNA inactivation as a host defense mechanism. Biol. Chem. Hoppe-Seyler 372, 557–564.

Duran-Reynals, F., 1942. The reciprocal infection of ducks and chickens with tumor inducing viruses. Cancer Res. 2, 343–369.

Emerman, M., 2000. Learning from lentiviruses. Nature Genet. 24, 8–9.

Fauci, A.S., 1996. Host factors and the pathogenesis of HIV-induced disease. Nature 384, 529–534.

Fazvaag, A., Dai, H.Y., Aarseth, H., Dalen, A.B., 1993. A low oncogenic variant of friend murine leukemia virus with strong immunosuppressive properties. Arch. Virol. 131, 265–275.

Fincham, V.J., Wyke, J.A., 1991. Differences between cellular integration sites of transcribed and nontranscribed Rous sarcoma proviruses. J. Virol. 65, 461–463.

Flanagan, J.R., Becker, K.G., Ennist, D.L., Gleason, S.L., Driggers, P.H., Levi, B.Z., Appella, E., Ozato, K., 1992. Cloning of a negative transcription factor that binds to the upstream conserved region of Moloney murine leukemia virus. Mol. Cell. Biol. 12, 38–44.

Gao, F., Yue, L., White, A.T., Pappas, P.G., Barchue, J., Hanson, A.P., Greene, B.M., Hahn, B.H., 1992. Human infection by genetically diverse SIV_{SM}-related HIV-2 in West Africa. Nature 358, 495–499.

Gao, F., Bailes, E., Robertson, D.L., Chen, Y., Rodenburg, C.M., Michael, S.F., Cummins, L.M., Arthur, L.O., Peeters, M., Shaw, G.M., Sharp, P.M., Hahn, B.H., 1999. Origin of HIV-1 in the chimpanzee. *Pan troglodytes troglodytes*. Nature 397, 436–441.

Hawley, R.G., Lieu, F.H., Fong, A.Z., Hawley, T.S., 1994. Versatile retroviral vectors for potential use in gene therapy. Gene Ther. 1, 136–138.

Hejnar, J., Plachá, J., Geryk, J., Machoň, O., Trejbalová, K., Guntaka, R.V., Svoboda, J., 1999. Inhibition of the Rous sarcoma virus long terminal repeat-driven transcription by in vitro methylation: different sensitivity in permissive chicken cells versus mammallian cells. Virology 255, 171–181.

Hirsch, V.M., Olmsted, R.A., Murphy-Corb, M., Purcell, R.H., Johnson, P.R., 1989. An African primate lentivirus (SIV_{sm}) closely related to HIV-2. Nature 339, 389–392.

Karakoz, I., Geryk, J., Svoboda, J., 1980. In vivo effect of three transformation-defective mutants of subgroup C avian sarcoma viruses. In: Zadrazil, S., Sponar, J. (Eds.), DNA – Recombination. Interactions and Repair. Pergamon Press, Oxford and New York, pp. 435–440.

Knoepfler, P.S., Eisenman, R.N., 1999. Sin meets NuRD and other tails of repression. Cell 99, 447–450.

Kucerova, L., Altanerova, V., Altaner, C., Boris-Lawrie, K., 1999. Bovine

leukemia virus structural gene vectors are immunogenic and lack pathogenicity in a rabbit model. J. Virol. 73, 8160–8166.
Levy, J.A., 1977. Murine xenotropic type C viruses III. Phenotypic mixing with avian leukosis and sarcoma viruses. Virology 77, 811–825.
Levy, J.A., 1978. Xenotropic type C viruses. Curr. Top. Microbiol. Immunol. 79, 111–213.
Levy, J.A., Oleszko, O., Dimpfl, J., Lau, D., Rigdon, R.H., Jones, J., Avery, R., 1982. Murine xenotropic type C viruses IV. Replication and pathogenesis in ducks. J. Gen. Virol. 61, 65–74.
Lorincz, M.C., Schübeler, D., Goeke, S.C., Walters, M., Groudine, M., Martin, D.I.K., 2000. Dynamic analysis of proviral induction and de novo methylation: implications for a histone deacetylase-independent, methylation density-dependent mechanism of transcriptional repression. Mol. Cell. Biol. 20, 842–850.
Machoň, O., Strnad, V., Hejnar, J., Geryk, J., Svoboda, J., 1998. Sp1 binding sites inserted into the Rous sarcoma virus long terminal repeat enhance LTR-driven gene expression. Gene 208, 73–82.
Macleod, D., Charlton, J., Mullins, J., Bird, A.P., 1994. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. Genes Dev. 8, 2282–2292.
Mazgareanu, S., Muller, J.G., Czub, S., Schimmer, S., Bredt, M., Czub, M., 1998. Suppression of rat bone marrow cells by Friend murine leukemia virus envelope proteins. Virology 242, 357–365.
Mummaneni, P., Bishop, P.L., Turker, M.S., 1993. A cis-acting element accounts for a conserved methylation pattern upstream of the mouse adenine phosphoribosyltransferase gene. J. Biol. Chem. 268, 552–558.
Nelson, D.T., Goodchild, N.L., Mager, D.L., 1996. Gain of Sp1 sites and loss of repressor sequences associated with a young, transcriptionally active subset of HERV-H endogenous long terminal repeats. Virology 220, 2213–2218.
Ng, H.-H., Bird, A., 1999. DNA methylation and chromatin modification. Curr. Opin. Genet. Dev. 9, 158–163.
Osborne, C.S., Pasceri, P., Singal, R., Sukonnik, T., Ginder, G.D., Ellis, J., 1999. Amelioration of retroviral vector silencing in locus control region β-globin-transgenic mice and transduced F9 embryonic cells. J. Virol. 73, 5490–5496.
Prince, V.E., Righy, P.W.J., 1991. Derivatives of Moloney murine sarcoma virus capable of being transcribed in embryonal carcinoma stem cells have gained a functional Sp1 binding site. J. Virol. 65, 1803–1811.
Razin, A., 1998. CpG methylation, chromatin structure and gene silencing – a three-way connection. EMBO J. 17, 4905–4908.
Rivella, S., Callegari, J.A., May, C., Tan, C.W., Sadelain, M., 2000. The cHS4 insulator increases the probability of retroviral expression at random chromosomal integration sites. J. Virol. 74, 4679–4687.
Ryndinch, A., Kadl, R., Geryk, J., Zoubak, S., Svoboda, J., Bernardi, G., 1991. The isopycnic compartmentalized integration of Rous sacoma virus sequences. Gene 106, 165–172.
Saha, K., Yuen, P.H., Wong, P.K.Y., 1994. Murine retrovirus-induced depletion of T cells is mediated through activation-induced death by apoptosis. J. Virol. 68, 2735–2740.
Schulze-Forster, K., Gotz, F., Wagner, H., Kroger, H., Simon, D., 1990. Transcription of HIV 1 is inhibited by DNA methylation. Biochem. Biophys. Res. Commun. 168, 141–147.
Sharp, P.M., Robertson, D.L., Hahn, B.H., 1995. Cross-species transmission and recombination of 'AIDS' viruses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 349, 41–47.
Sjottem, E., Anderssen, S., Johansen, T., 1996. The promoter activity of long terminal repeats of the HERV-H family of human retrovirus-like elements is critically dependent on Sp1 family proteins interacting with a GC/GT box located immediately 3′ to the TATA box. J. Virol. 70, 188–198.
Smith, R.E., Schmidt, E.V., 1982. Induction of anemia by avian leukosis viruses of five subgroups. Virology 117, 516–518.
Stepanets, V., Vernerová, Z., Vilhelmová, M., Geryk, J., Hejnar, J., Weichsel, F.F., Svoboda, J., 2000. Intraembryonic avian leukosis virus subgroup C (ALV-C) inoculation producing wasting disease in ducks soon after hatching (submitted).
Stewart, C.L., Stuhlmann, H., Jahner, D., Jaenisch, R., 1982. De novo methylation, expression, and infectivity of retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 79, 4098–4102.
Svoboda, J., 1986. Rous sarcoma virus. Intervirology 26, 1–60.
Svoboda, J., 1998. Molecular biology of cell non-permissiveness to retroviruses. Has the time come? Gene 206, 153–163.
Trejbalová, K., Gebhard, K., Vernerová, Z., Dušek, L., Geryk, J., Hejnar, J., Haase, A.T., Svoboda, J., 1999. Proviral load and expression of avian leukosis viruses subgroup C in long-term persistently infected heterologous hosts (ducks). Arch. Virol. 144, 1779–1807.
Weiss, R.A. Wrangham, R.W., 1999. From *pan* to pandemic. Nature 385–386.
Wolffe, A.P., Matzke, M.A., 1999. Epigenetics: regulation through repression. Science 286, 481–486.
Wyatt, C.R., Wingett, D., White, J.S., Buck, C.D., Knowles, D., Reeves, R., Magnuson, N.S., 1989. Persistent infection of rabbits with bovine leukemia virus associated with development of immune dysfunction. J. Virol., 63, 4498–4506.
Yang, W.M., Inouye, C., Zeng, Y., Bearss, D., Seto, E., 1996. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator RPD3. Proc. Natl. Acad. Sci. USA 93, 12845–12850.
Yoder, J.A., Walsh, C.P., Bestor, T.H., 1997. Cytosine methylation and the ecology of intragenomic parasites. Trends Genet. 13, 335–340.
Young, W.-B., Lindberg, G.L., Link Jr, C.J., 2000. DNA methylation of helper virus increases genetic instability of retroviral producer cells. J. Virol. 74, 3177–3187.

TAB 7

This material may be protected by Copyright law (Title 17 U.S. Code)

Clin Genet 2001: 59: 17–24
Printed in Ireland. All rights reserved

Copyright © Munksgaard 2001
CLINICAL GENETICS
ISSN 0009-9163

Developmental Biology: Frontiers for Clinical Genetics

Section Editor: Roderick R McInnes
e-mail: mcinnes@sickkids.on.ca

The beta-globin locus control region versus gene therapy vectors: a struggle for expression

Ellis J, Pannell D. The beta-globin locus control region versus gene therapy vectors: a struggle for expression.
Clin Genet 2001: 59: 17–24. © Munksgaard, 2001

Developmental control of gene expression has a major impact on the design of β-globin retrovirus vectors for hematopoietic stem cell gene therapy of β-thalassemia. It is obvious that the endogenous locus control region (LCR) elements that drive β-globin gene expression in transgenic mice must be included in these vectors. However, the specific elements to use are not clear and require an understanding of LCR action. Moreover, retrovirus vectors contain silencer elements that function in stem cells and are dominant to LCR function. Recent studies on LCRβ-globin transgenes and retrovirus silencing suggest ways to overcome this silencing effect after transfer into stem cells and carefully designed lentivirus vectors have exciting therapeutic benefit in animal models of β-thalassemia. By building on 15 years of development, LCRβ-globin vectors are now being tested in preclinical animal models and may ultimately lead to the long-sought cure for this genetic disease.

J Ellis and D Pannell

Developmental Biology Program, Hospital for Sick Children, 555 University Avenue, Toronto, ON, and Department of Molecular and Medical Genetics, University of Toronto, Canada.

Key words: chromatin – gene silencing – gene therapy – LCR

Corresponding author: J Ellis, Developmental Biology Program, Hospital for Sick Children, 555 University Avenue, Toronto; ON, Canada M5G 1X8. Tel: +416 813 7295; fax: +416 813 8883; e-mail: jellis@sickkids.on.ca

Received 12 October 2000, revised and accepted for publication 16 October 2000

β-thalassemia is caused by genetic defects that reduce β-globin protein levels (1). The resulting imbalance of β-globin to α-globin chains results in a severe anemia that is generally treated by repeated blood transfusions. A consequence of such regular blood transfusion is the increased risk of exposure to virus-infected blood supplies, and eventually leads to iron overload that can cause organ failure (2). Present iron chelation treatments are controversial (3), and the only cure available is bone marrow transplantation from a matched sibling if available. An attractive alternative is to perform gene therapy to deliver a human β-globin gene into hematopoietic stem cells (HSC) from the patient (4, 5).

For β-globin gene therapy to be successful, it is essential that the transferred gene be expressed to the correct level. This goal requires a detailed knowledge of the mechanism and *cis*-acting sequences that control β-globin expression during development. Surprisingly, inclusion of the appropriate regulatory elements may not be sufficient to obtain therapeutic levels during gene therapy as the vectors used to deliver the gene are frequently silenced in transduced stem cells. Hence, a better understanding of the mechanism of vector silencing in stem cells is also required. Here, we review the regulatory elements that control β-globin gene expression during development and their use in β-globin retrovirus vectors, outline the evidence that retrovirus and lentivirus vectors are silenced in stem cells and potential means to overcome this silencing, and we conclude with preclinical animal models to test promising β-globin gene therapy vectors.

β-globin gene expression during development

The β-globin gene is part of a cluster of highly related globin genes located on Chr 11p15 in humans (6). These genes are arranged in the same order as they are expressed during development (Fig. 1). The ε-globin gene is expressed in the blood islands of the yolk sac, the site of hemato-

poiesis then switches to the fetal liver where the γ-globin genes are expressed, and shortly after birth hematopoiesis switches to the bone marrow where the δ- and β-globin genes are expressed to very low and high levels, respectively. Many mutations that cause β-thalassemia have been described and some have been informative with regard to the sequences and molecular mechanisms that control globin gene switching. For example, it is clear that point mutations in the γ-globin promoters can enhance expression of these genes in adults causing hereditary persistence of fetal hemoglobin (HPFH). Deletion of upstream sequences that include the locus control region (LCR) results in a lack of expression from the still intact globin genes. These data demonstrate that promoter sequences and the LCR are important control elements for globin expression.

The β-globin locus control region

The LCR is composed of at least four DNaseI hypersensitive sites (HS) located upstream of the locus (Fig. 1) (7–9). The presence of HS indicates that *trans*-acting factors are binding to these regions and displacing or destabilizing nucleosomes. Nucleosomes are the basic units of chromatin and condense DNA around an octomer of the histone proteins H2A, H2B, H3 and H4 (10). Expressed genes are located an 'open' chromatin that is more accessible to *trans*-acting factors and in general contain nucleosomes with highly acetylated histones. In contrast, 'closed' chromatin generally has deacetylated histones and is bound by the linker histone H1, is less accessible to DNA binding factors and genes in these regions are not expressed. Chromatin structure is modulated using many chromatin remodelling complexes (11, 12).

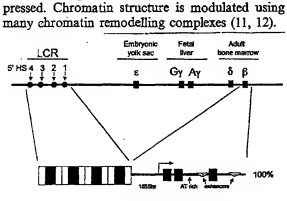


Fig. 1. Gene structure and expression pattern of the human β-globin locus. The locus control region (LCR) is composed of at least 4 DNaseI hypersensitive sites (HS) located upstream of the cluster. *Trans*-acting factors (red circles) bind to the HS. In transgenic mice, LCR activity directs copy number-dependent, position-independent transgene expression. Full levels of β-globin transgene expression are obtained in the presence of all four HS including their core elements (red boxes).

The mechanism by which the LCR controls β-globin gene expression has been extensively studied, primarily using transgenic and knockout mice (6). It is important to realize that these two assays manipulate the genes in quite different ways and the results are not always complementary or in agreement with each other. Transgenic mice contain the human β-globin gene transferred into novel or ectopic integration sites; whereas, knockout mice manipulate the endogenous native locus in the mouse. In transgenic mice, human β-globin transgenes are silent at most integration sites or transcribed to about 1% of the endogenous mouse βmajor level. In contrast, addition of the LCR, including all four HS to the β-globin transgene results in expression to about 100% levels at all integration sites, and expression is copy number-dependent (Fig. 1) (7, 13, 14). This copy number-dependent, position-independent transgene expression is unusual and is the defining feature of LCR activity. Further investigation demonstrated that individual HS2, HS3 and HS4 elements and their smaller 'cores' of approximately 200–300 bp, also direct copy number-dependent transgene expression but to lower levels (10–25%) (15–20).

The LCR is often referred to as an enhancer, but does not have classic enhancer activity because it does not function equally well in either orientation (21). Rather, it appears that complete LCR activity requires all four HS (22), and these have some distinct roles. For example, HS3 can activate β-globin transgenes at all single copy integration sites where it establishes open chromatin and remodels chromatin on the promoter to permit expression (13). In contrast, although HS2 has strong enhancer activity in transient transfection studies (23), it is unable to direct expression in single copy transgenic mice (24). These data suggest that at ectopic sites, the HS function together as a unit, making the LCR sufficient to open chromatin and enhance full expression of β-globin transgenes.

Open chromatin is likely to be established by the binding of erythroid *trans*-acting factors that recruit chromatin remodelling complexes (25–31), as has been described for histone acetylation changes on active human β-globin genes (32). This open chromatin may not extend throughout transgenes containing the entire human LCRβ-globin cluster, as different domains that correlate with the presence of low level intergenic transcription have been described during globin switching in mice (33).

Two models of LCR activity

The transgenic mouse data have largely been interpreted as supporting a holocomplex model of LCR

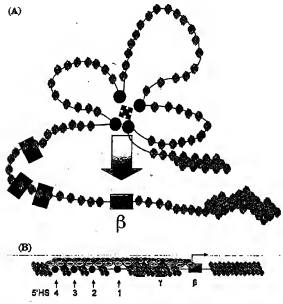


Fig. 2. Two models of LCR activity. A) The holocomplex model suggests that factors (red circles) bound to the HS interact with each other by DNA looping (arrows) to form a single LCR holocomplex that loops to activate expression from the appropriate globin promoter. B) The linking model suggests that the LCR serves to ensure that factors (coloured circles) are bound throughout the locus. Nucleosomes in 'open' chromatin are represented as dispersed green circles, 'closed' chromatin as condensed solenoids of green circles.

action at ectopic sites (34). Once open chromatin has been established, it is proposed that each of the HS then interacts with each other by DNA looping mediated via the bound factors to form an LCR holocomplex (Fig. 2A). The holocomplex would then interact with a single globin gene in the cluster, and switching during development would be accomplished by stage-specific silencer elements associated with the ε- and γ-globin genes. Although there is no direct physical evidence for DNA looping between the LCR and the globin promoters *in vivo*, the holocomplex model is supported by findings that only a single globin gene is transcribed at a time in transgenic mice containing the whole human LCRβ-globin cluster (35), and that the LCR preferentially activates genes closest to it (36). Deletions that remove only a single HS 'core' element drastically reduce transgene expression in comparison to those that delete an entire HS fragment (37, 38), indicating that removal of a core creates a defective holocomplex (39).

Quite different conclusions have been arrived at using knockout technology on the mouse β-globin locus (40). Deletions of individual or all the HS in the endogenous locus do not alter chromatin structure and have relatively minor effects on expression of the globin genes (41, 42). These data suggest that the LCR is not required for chromatin opening at the endogenous mouse β-globin locus, and suggest that more distant elements control chromatin structure (43). A linking model that does not invoke DNA looping has been proposed to explain the knockout results (Fig. 2B) (44). In this model, the function of the LCR is to enhance β-globin expression by ensuring that factors are bound at intervals across the cluster and that the gene is localized to the right nuclear compartment (32). The linking model is not consistent with the ability of the LCR to open chromatin at ectopic transgene sites, but the holocomplex model cannot easily explain the effect of LCR deletions in the mouse β-globin locus. As described below, the two models are not necessarily mutually exclusive and may be strengthened by being merged. For the purpose of gene therapy where globin expression cassettes delivered by viral vectors must express at ectopic sites, it will be important to design the cassettes based on transgene constructs that express to high levels at single copy integration sites.

LCRβ-globin expression cassettes for gene therapy

To express therapeutic levels of β-globin from gene therapy cassettes, full expression levels should be obtained from a single copy integration in order to convert a null thalassemia into an asymptomatic carrier state. Initial β-globin gene therapy cassettes were designed to be as small as possible to facilitate gene transfer and used either cDNA or genomic β-globin fragments controlled by minimal promoters (45–47). Addition of small HS core elements improves expression in tissue culture experiments, but were largely disappointing when transferred into mouse bone marrow cells (48, 49). Recent single copy transgenic mouse experiments now demonstrate that full expression by the LCR requires all four HS and specific elements within the β-globin gene including the −1555 bp promoter and the 3′ enhancer (Fig. 3 top) (50). As this 8.8-kb cassette is too large for conventional retrovirus vectors, smaller constructs that express highly are required for retrovirus delivery. A very promising new 3.9-kb cassette expresses γ-globin mRNA to 70% levels in single copy transgenic mice, using β-globin promoter, intron 2 and 3′ enhancer elements (Fig. 3 centre) (51). This construct, or others shown to function effectively in transgenic mice, may ultimately prove to be best suited for gene therapy and take advantage of the anti-sickling properties of γ-globin. In addition, these data demonstrate that the LCR must functionally interact with more than just the promoter. The simplest

interpretation merges the holocomplex and linking models by suggesting that the LCR loops to interact with the promoter, but factors must also be bound throughout the gene.

Retrovirus and lentivirus vectors

Retrovirus vectors have been the method of choice for delivering β-globin genes into hematopoietic stem cells because they stably integrate at single copy into the genome (52). In practice, it has been very difficult to obtain high titer β-globin retrovirus vectors due to instability of the LCR elements and globin intron 2 sequences (48, 49). To stabilize transmission of intact β-globin genes it is necessary to use only certain combinations of HS sites and to delete an AT-rich region in intron 2. Although these modifications permit generation of high titer retrovirus, it is now apparent that the deleted AT-rich sequence is required for high level expression (51).

A limitation of retrovirus vectors is that they integrate only into cycling cells, and the target HSC for β-globin gene therapy tend to be non-cycling. In this regard, lentivirus vectors based on HIV-1 (53) are far superior as they integrate into non-cycling cells and contain the RRE element that is bound by the Rev protein to stabilize the virus genomic RNA (52). One exciting report of a β-globin lentivirus vector demonstrates that it is possible to transmit large LCR fragments coupled to a β-globin cassette with a small promoter, AT-rich deleted intron 2 and no 3′ enhancer (Fig. 3 bottom). Therapeutic levels of β-globin mRNA and protein were shown in transduced bone marrow in a mouse model of β-thalassemia (54).

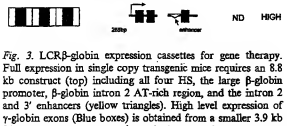


Fig. 3. LCRβ-globin expression cassettes for gene therapy. Full expression in single copy transgenic mice requires an 8.8 kb construct (top) including all four HS, the large β-globin promoter, β-globin intron 2 AT-rich region, and the intron 2 and 3′ enhancers (yellow triangles). High level expression of γ-globin exons (Blue boxes) is obtained from a smaller 3.9 kb β/γ-globin hybrid cassette (centre). As both of these cassettes are not stable in retrovirus vectors, cassettes with improved stability (bottom) have deletions of deleterious sequences.

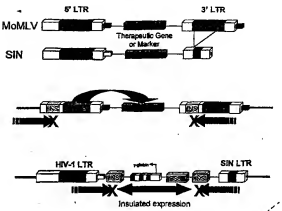


Fig. 4. Retrovirus silencing in stem cells. Silencer elements in MoMLV vectors are primarily located in the LTRs (top), and silencing can be ameliorated but not eliminated by SIN mutations in the 3′LTR. Insulator elements (INS) in the LTR (centre) block silencing by position effects (red arrows with X) in MEL cells but not retrovirus silencing in stem cells (large red arrow). The ideal vector (bottom) uses a SIN lentivirus vector, insulator elements flanking a small HS3β/γ-globin expression cassette to prevent silencing (green arrow), and no marker gene.

Silencing of retrovirus vectors in stem cells

Once the β-globin expression cassette has been transduced into an HSC, it must express to appropriate levels in erythroid cells produced from the HSC. One concern that has severely limited clinical gene therapy is that retrovirus-transduced genes are often silenced in HSC (52, 55–57). It has long been known that retroviruses are silenced in embryonic stem cells and in transgenic mice (58, 59). Moreover, it has been shown that retrovirus and lentivirus sequences silence LCRβ-globin transgenes in mice (60, 61), and that the transgenes fail to open chromatin (60). These data demonstrate that the vectors have silencer elements, and suggest that they are recognized by specific factors that are restricted to stem cells in mammals.

There are several possible means to overcome vector silencing in stem cells. First, the silencer elements can be defined in the retrovirus or lentivirus vectors and, if not required for virus replication, can be removed. As the silencers tend to be located in the viral long terminal repeats (LTRs) that control virus transcription (Fig. 4 top), mutations produce self-inactivating (SIN) virus that is inherently safer. Several SIN retrovirus vectors have been developed from Moloney murine leukemia virus that express to higher levels in ES cells and transgenic mice (62–65). Nevertheless, none of the SIN vectors express at all integration sites in stem cells. To overcome this residual silencing in the SIN vectors, it may be possible to

incorporate insulator elements to block the spread of silencing from surrounding retrovirus sequences (66). The chicken β-globin cHS4 element is an insulator (66, 67) and has been placed in the LTR of retrovirus vectors to block position effects in murine erythroleukemia (MEL) cells (Fig. 4 centre) (68, 69). However, an insulator in the LTR does not shield internal genes from silencing established on the retrovirus sequences in ES cells (68). A better construct design would position insulators on both sides of the internal gene rather than in the LTR (Fig. 4 centre).

Mechanism of retrovirus vector silencing

An understanding of the mechanism of retrovirus silencing may permit interventions that prevent its establishment or maintenance (70, 71). Many groups have correlated *de novo* cytosine methylation (72) of CpG dinucleotides in retroviruses with the silenced state (58, 59, 68, 73, 74), and expression can be reactivated to a low level using the methylation inhibitor 5 AzaCytidine (5AzaC) (75). However, this indirect evidence does not address whether methylation is a cause of silencing or a consequence. Recent direct evidence that retrovirus silencing is independent of *de novo* methylase function has been obtained using dnmt3 knockout ES cells and transgenic *Drosophila* that have no methylase activity (60). As chromatin of retrovirus-silenced LCRβ-globin transgenes in mice is inaccessible to DNaseI and marked by deacetylated histone H3 and bound linker H1 (60), it appears that chromatin modifications play a role in retrovirus silencing. Attempts to relieve silencing may, therefore, require inhibitors of histone deacetylases (HDAC) or of H1 binding.

Silencing is often established by one pathway and maintained by another. Time course experiments in ES cells demonstrate that most retroviruses are silenced within 2 days but some integration sites escape complete silencing and express to low levels (60, 76, 77). As methylation is not detectable by 2 days post-infection, methylation is likely to be a consequence of, or secondary step in, retrovirus silencing. Most experiments have focussed on the subset of infected cells that initially express. Over time, these are gradually silenced in a process known as extinction. In infected MEL cells, extinction can be reversed early in the process using the HDAC inhibitor Trichostatin A (TSA) (78). However, the methylation inhibitor 5 AzaC is required in addition to TSA to overcome extinction at later time points. These data suggest that methylation is an important secondary or associated step in extinction of virus expression in mature cell types. The ability of TSA or 5 AzaC to activate expression in the majority of transduced cells that are completely silenced from the outset has not been rigourously tested to date. In summary, efforts to prevent retrovirus silencing using methylation and HDAC inhibitors hold promise, but require more knowledge of the mechanism and demonstration of their utility in silenced stem cell populations prior to extinction.

An ideal LCRβ-globin lentivirus vector

The ideal β-globin gene therapy vector should stably integrate into an HSC at high efficiency and be expressed to near endogenous levels at single copy. To accomplish this goal, lentivirus vectors have a clear advantage in their ability to infect non-cycling stem cells and stably transmit large LCRβ-globin expression cassettes. The best existing LCRβ-globin lentivirus vector has these features (54), with the additional advantage of omitting a selectable marker gene. Although marker genes are convenient for determining transduction frequencies, most are derived from non-mammalian sources and may themselves be subject to gene silencing effects. However, the vector can be optimized further to direct expression at all integration sites and for vector safety (Fig. 4 bottom). First, a SIN version of the lentiviral vectors must now be used with third generation packaging systems designed to prevent recombination events that generate replication-competent HIV-1 virus (79). The SIN lentivirus vector will not only prevent rescue and spread of the vector by any helper virus, but may also improve expression of the LCRβ-globin cassette (60). Second, the LCRβ-globin cassette should be flanked by insulator elements that are known to block silencing. In practise, this may require a different insulator on one side than the other to avoid recombination events that delete the LCRβ-globin cassette. Finally, the LCRβ-globin cassette should express highly at single copy in transgenic mice. Such constructs use HS3 coupled to a large β-globin promoter, the β-globin intron 2 including both the AT-rich region and enhancer and the 3′ enhancer (51). Use of hybrid genes permits expression of anti-sickling γ- or δ-globin coding sequences instead of β-globin exons (51, 80). A combination of these components should create a lentivirus vector that is safe and expresses therapeutic levels of globin.

Preclinical models of β-thalassemia

Promising β-globin gene therapy vectors have been tested primarily in MEL cells, or in infected mouse

bone marrow. Expression in erythroid cells derived from primitive progenitor cells can be assayed using CFU-S assays in which spleen colonies are formed after 12 days *in vivo*. Expression in cells derived from an infected HSC must be assayed after long-term repopulation assays, followed by secondary transplantation. To show therapeutic efficacy, these long-term studies should be performed first on mouse models of β-thalassemia and sickle cell anemia. A variety of these models have been created by gene targeting and transgenic technology (81–83). These mouse models can be corrected by expression of γ-globin transgenes, and it has been shown that an LCRβ-globin lentivirus vector can express β-globin mRNA and protein in these models with therapeutic benefit (54).

Despite success with the mouse models, it has proved much more difficult to transduce human HSC than murine HSC (84). Fortunately, another preclinical model of β-globin gene therapy into human stem cells is available (Fig. 5). The NOD-Scid mouse is severely immunocompromized and fails to reject human bone marrow transplants (85). Transplanted human HSC home to the mouse bone marrow where they are supported by the hematopoietic microenvironment. Human bone marrow cells from β-thalassemia and sickle cell patients have been shown to repopulate the bone marrow of these mice and generate human red cells that mimic the disease. This system is very well suited to test expression from LCRβ-globin lentivirus vectors in human stem cells. To this end, patient bone marrow or sorted HSC would be infected with the lentivector *in vitro* prior to *in vivo* assay in the NOD/Scid mice. Hence, long-term expression in human stem cells can be assayed *in vivo* without exposing patients to experimental lentivirus vectors.

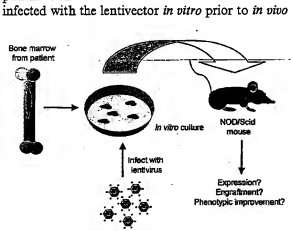


Fig. 5. The NOD-Scid preclinical mouse model of human stem cell gene therapy for β-thalassemia or sickle cell anemia. Patient bone marrow stem cells are infected with the β-globin lentivirus vector *in vitro*, prior to *in vivo* engraftment into immunocompromised NOD-Scid mice. Gene transfer into engrafted human stem cells, β-globin transgene expression, and phenotypic improvement can then be monitored without exposing patients to experimental lentivirus vectors.

Future prospects

The first retrovirus vectors for β-globin gene therapy were designed over 15 years ago, and through slow careful research many obstacles were discovered and gradually surmounted. The success of a well-designed LCRβ-globin lentivirus vector in correcting a mouse β-thalassemia model is a milestone in this process that can now be completed through incremental improvements to vector expression and safety. Ultimately, validation of β-globin gene therapy in human stem cells using the NOD-Scid preclinical model will justify clinical trials of this exciting potential cure for hemoglobinopathies.

Acknowledgements

We acknowledge the Medical Research Council (MRC) of Canada grant support (to JE) for our β-globin expression and retrovirus silencing research and Hospital for Sick Children Foundation, OGS and MRC Doctoral Research Awards to DP. We regret that many fine studies relevant to this review have been omitted due to space constraints.

References

1. Weatherall DJ. The Thalassemias. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001: 183–226.
2. Olivieri NF. The beta-thalassemias. N Engl J Med 1999: 341 (2): 99–109.
3. Olivieri NF et al. Long-term safety and effectiveness of iron-chelation therapy with deferiprone for thalassemia major. N Engl J Med 1998: 339 (7): 417–423.
4. Persons DA, Nienhuis AW. Gene therapy for the hemoglobin disorders: past, present, and future. Proc Natl Acad Sci USA 2000: 97 (10): 5022–5024.
5. Sorrentino BP, Nienhuis AW. Gene therapy for hematopoietic diseases. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. W.B. Saunders: Philadelphia, 2001: 969–1003.
6. Stamatoyannopoulos G, Grosveld F. Hemoglobin switching. In: Stamatoyannopoulos G et al., eds. The Molecular Basis of Blood Diseases. Philadelphia: W.B. Saunders, 2001: 135–182.
7. Grosveld F et al. Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 1987: 51 (6): 975–985.
8. Forrester WC et al. Evidence for a locus activation region: the formation of developmentally stable hypersensitive sites in globin-expressing hybrids. Nucl Acid Res 1987: 15 (24): 10159–10177.
9. Tuan D, London IM. Mapping of DNase I-hypersensitive sites in the upstream DNA of human embryonic epsilon-globin gene in K562 leukemia cells. Proc Nalt Acad Sci USA 1984: 81 (9): 2718–2722.
10. Felsenfeld G. Chromatin unfolds. Cell 1996: 86 (1): 13–19.

11. Vignali M et al. ATP-dependent chromatin-remodeling complexes. Mol Cell Biol 2000: 20 (6): 1899–1910.
12. Kingston RE, Narlikar GJ. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. Genes Dev 1999: 13 (18): 2339–2352.
13. Ellis J et al. A dominant chromatin-opening activity in 5' hypersensitive site 3 of the human beta-globin locus control region. EMBO J 1996: 15 (3): 562–568.
14. Talbot D et al. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. Nature 1989: 338 (6213): 352–355.
15. Philipsen S et al. The beta-globin dominant control region: hypersensitive site 2. EMBO J 1990: 9 (7): 2159–2167.
16. Philipsen S, Pruzina S, Grosveld F. The minimal requirements for activity in transgenic mice of hypersensitive site 3 of the beta globin locus control region. EMBO J 1993: 12 (3): 1077–1085.
17. Pruzina S et al. Hypersensitive site 4 of the human beta globin locus control region. Nucl Acid Res 1991: 19 (7): 1413–1419.
18. Talbot D et al. Detailed analysis of the site 3 region of the human beta-globin dominant control region. EMBO J 1990: 9 (7): 2169–2177.
19. Talbot D, Grosveld F. The 5'HS2 of the globin locus control region enhances transcription through the interaction of a multimeric complex binding at two functionally distinct NF-E2 binding sites. EMBO J 1991: 10 (6): 1391–1398.
20. Caterina JJ et al. Human beta-globin locus control region: analysis of the 5' DNase I hypersensitive site HS 2 in transgenic mice. Proc Natl Acad Sci USA 1991: 88 (5): 1626–1630.
21. Tanimoto K et al. Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. Nature 1999: 398 (6725): 344–348.
22. Fraser P et al. Each hypersensitive site of the human beta-globin locus control region confers a different developmental pattern of expression on the globin genes. Genes Dev 1993: 7 (1): 106–113.
23. Tuan DY et al. An erythroid-specific, developmental-stage-independent enhancer far upstream of the human 'beta-like globin' genes. Proc Natl Acad Sci USA 1989: 86 (8): 2554–2558.
24. Ellis J et al. Synthetic human beta-globin 5'HS2 constructs function as locus control regions only in multicopy transgene concatamers. EMBO J 1993: 12 (1): 127–134.
25. Blobel GA et al. CREB-binding protein cooperates with transcription factor GATA-1 and is required for erythroid differentiation. Proc Natl Acad Sci USA 1998: 95 (5): 2061–2066.
26. Boyes J et al. Perturbation of nucleosome structure by the erythroid transcription factor GATA-1. J Mol Biol 1998: 279 (3): 529–544.
27. Boyes J et al. Regulation of activity of the transcription factor GATA-1 by acetylation. Nature 1998: 396 (6711): 594–598.
28. Armstrong JA, Emerson BM. NF-E2 disrupts chromatin structure at human beta-globin locus control region hypersensitive site 2 *in vitro*. Mol Cell Biol 1996: 16 (10): 5634–5644.
29. Armstrong JA, Bieker JJ, Emerson BM. A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF *in vitro*. Cell 1998: 95 (1): 93–104.
30. Gong QH, McDowell JC, Dean A. Essential role of NF-E2 in remodeling of chromatin structure and transcriptional activation of the epsilon-globin gene *in vivo* by 5' hypersensitive site 2 of the beta-globin locus control region. Mol Cell Biol 1996: 16 (11): 6055–6064.
31. O'Neill D et al. Tissue-specific and developmental stage-specific DNA binding by a mammalian SWI/SNF complex associated with human fetal-to-adult globin gene switching. Proc Natl Acad Sci USA 1999: 96 (2): 349–354.
32. Schubeler D et al. Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human beta-globin locus. Genes Dev 2000: 14 (8): 940–950.
33. Gribnau J et al. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol Cell 2000: 5 (2): 377–386.
34. Grosveld F. Activation by locus control regions? Curr Opin Genet Dev 1999: 9 (2): 152–157.
35. Wijgerde M, Grosveld F, Fraser P. Transcription complex stability and chromatin dynamics *in vivo*. Nature 1995: 377 (6546): 209–213.
36. Dillon N et al. The effect of distance on long-range chromatin interactions. Mol Cell 1997: 1 (1): 131–139.
37. Bungert J et al. Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. Genes Dev 1995: 9 (24): 3083–3096.
38. Peterson KR et al. Effect of deletion of 5'HS3 or 5'HS2 of the human beta-globin locus control region on the developmental regulation of globin gene expression in beta-globin locus yeast artificial chromosome transgenic mice. Proc Natl Acad Sci USA 1996: 93 (13): 6605–6609.
39. Wood WG. The complexities of β globin gene regulation. Trends Genet 1996: 12 (6): 204–206.
40. Higgs DR. Do LCRs open chromatin domains? Cell 1998: 95 (3): 299–302.
41. Reik A et al. The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. Mol Cell Biol 1998: 18 (10): 5992–6000.
42. Epner E et al. The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. Mol Cell 1998: 2 (4): 447–455.
43. Bulger M et al. Conservation of sequence and structure flanking the mouse and human beta-globin loci: the beta-globin genes are embedded within an array of odorant receptor genes. Proc Natl Acad Sci USA 1999: 96 (9): 5129–5134.
44. Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. Genes Dev 1999: 13 (19): 2465–2477.
45. Cone RD et al. Regulated expression of a complete human beta-globin gene encoded by a transmissible retrovirus vector. Mol Cell Biol 1987: 7 (2): 887–897.
46. Dzierzak EA, Papayannopoulou T, Mulligan RC. Lineage-specific expression of a human beta-globin gene in murine bone marrow transplant recipients reconstituted with retrovirus-transduced stem cells. Nature 1988: 331 (6151): 35–41.
47. Karlsson S et al. Expression of the human beta-globin gene following retroviral-mediated transfer into multipotential hematopoietic progenitors of mice. Proc Natl Acad Sci USA 1988: 85 (16): 6062–6066.
48. Sadelain M et al. Generation of a high-titer retroviral vector capable of expressing high levels of the human beta-globin gene. Proc Natl Acad Sci USA 1995: 92 (15): 6728–6732.
49. Leboulch P et al. Mutagenesis of retroviral vectors transducing human beta-globin gene and beta-globin locus control region derivatives results in stable transmission of an

TAB 8

Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 2567–2571, March 1994
Medical Sciences

Lack of expression from a retroviral vector after transduction of murine hematopoietic stem cells is associated with methylation *in vivo*

(gene therapy/gene expression/bone marrow/long terminal repeat)

PIA-MARIA CHALLITA* AND DONALD B. KOHN*†‡

Departments of *Microbiology and †Pediatrics, University of Southern California School of Medicine, Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, Los Angeles, CA 90027

Communicated by Elizabeth F. Neufeld, November 18, 1993 (received for review July 1, 1993)

ABSTRACT **We describe studies of gene transfer and expression of the human glucocerebrosidase cDNA by a Moloney murine leukemia virus (MoMuLV)-based retroviral vector in a murine gene transfer/bone marrow transplant (BMT) model. Pluripotent hematopoietic stem cells (HSCs) were assayed as the colony-forming units, spleen (CFU-S) generated after serial transplantation. Transcriptional expression from the MoMuLV long-terminal repeat (LTR) was detected at a high level in the primary (1°) CFU-S and tissues of reconstituted BMT recipients. However, we observed transcriptional inactivity of the proviral MoMuLV-LTR in >90% of the secondary (2°) CFU-S and in 100% of the tertiary (3°) CFU-S examined. We have compared the methylation status of the provirus in the 1° CFU-S, which show strong vector expression, to that of the transcriptionally inactive provirus in the 2° and 3° CFU-S by Southern blot analysis using the methylation-sensitive restriction enzyme *Sma* I. The studies demonstrated a 3- to 4-fold increase in methylation of the *Sma* I site in the proviral LTR of 2° and 3° CFU-S compared to the transcriptionally active 1° CFU-S. These observations may have important implications for future clinical applications of retroviral-mediated gene transfer into HSCs, where persistent gene expression would be needed for an enduring therapeutic effect.**

Gene therapy via bone marrow cells is a promising technique for treatment of a wide variety of human diseases, including genetic disorders, cancer, and AIDS. Effective long-term bone marrow gene therapy requires the fulfillment of two main criteria. The exogenous gene should be introduced into a high percentage of long-lived pluripotent hematopoietic stem cells (HSCs). Subsequently, the introduced gene should be persistently expressed in the mature hematopoietic progeny cells of the stem cell, thereby maintaining the effects of gene therapy for the lifetime of the individual. Although Moloney murine leukemia virus (MoMuLV)-based retroviral vectors are currently the most efficient vehicles for gene transfer into a variety of cell types including HSCs (reviewed in ref. 1), the long-term *in vivo* expression from the viral promoter/enhancer elements has been unsatisfactory. Lack of gene expression from the 5′ MoMuLV long-terminal repeat (LTR) has been observed in several systems including primary fibroblasts (2) and hematopoietic cells (3, 4). Previous studies by our laboratory, using a retroviral vector in which a normal human glucocerebrosidase (GC) cDNA is controlled by the enhancer/promoter of the 5′ MoMuLV-LTR, demonstrated a high rate of lack of expression in cells derived from HSCs (5).

Methylation of cytosine residues has been shown to be associated with suppression of gene expression and, in certain circumstances, with the silencing of viral control elements (6). The MoMuLV-LTR is completely inactive in embryonic stem and embryonic carcinoma cell lines, and the inactivity is accompanied by *de novo* methylation of the proviral sequences (7, 8). Moreover, methylation has been detected in association with the MoMuLV-LTR transcriptional inactivity in fibroblasts *in vitro* (9) and *in vivo* (2).

In this study, we investigated long-term *in vivo* expression from the MoMuLV-LTR by transduction of murine bone marrow cells with a MoMuLV-based retroviral vector and serial bone marrow transplantation (BMT) into lethally irradiated recipient mice. We document a high rate of expression failure associated with methylation of the vector LTR in the secondary (2°) and tertiary (3°) colony-forming units, spleen (CFU-S).

MATERIALS AND METHODS

Retroviral Vector. The G2 retroviral vector and its corresponding high-titer amphotropic PA317 packaging cell clone have been described (5). G2 consists of the LTR from the N2 vector flanking the human GC cDNA. The packaging cell line clone used in the experiments was negative for helper virus production assayed by testing for transfer of the amphotropic *env* gene into 3T3 fibroblasts through PCR analysis (10).

Transduction of Murine Bone Marrow Cells. Donor bone marrow cells were harvested from male C57BL/6J mice (Charles River Breeding Laboratories), prestimulated in the presence of growth factors, and cocultivated over vector producing fibroblasts according to the methods described by Weinthal *et al.* (5). The growth factors used for the prestimulation were 200 units of murine interleukin 3 (IL-3) per ml (Biosource, Camarillo, CA), 100 units of human IL-6 per ml (Amgen), 200 units of human IL-1α per ml (Immunex) and 50 ng of mast cell growth factor per ml (or c-kit ligand; Immunex).

BMT and Sample Collection. Recipient female C57BL/6J mice (8–12 weeks old) were irradiated with two split doses of 600 and 450 cGy 24 hr apart. Transduced bone marrow cells were injected into the tail vein of the irradiated mice at 1×10^6 cells per mouse for isolation of CFU-S or $2–4 \times 10^6$ cells for long-term reconstitution. Twelve days after BMT, two to four mice transplanted with 1×10^6 bone marrow cells were sacrificed. Well-defined, individual primary (1°) CFU-S were isolated and divided evenly into two portions, one for DNA

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MoMuLV, Moloney murine leukemia virus; LTR, long terminal repeat; HSC, hematopoietic stem cell; GC, glucocerebrosidase; CFU-S, colony forming unit, spleen; BMT, bone marrow transplantation; 1°, primary; 2°, secondary; 3°, tertiary.

‡To whom reprint requests should be addressed at: Division of Research Immunology/Bone Marrow Transplantation, Childrens Hospital Los Angeles, 4650 Sunset Boulevard, Los Angeles, CA 90027.

Table 1. Expression of G2 in the mouse model of gene transfer/BMT

	Exp. 1		Exp. 2		Exp. 3		Total	
	DNA	RNA	DNA	RNA	DNA	RNA	DNA	RNA
1° CFU-S	9/9	5/5	12/12	6/6	8/8	8/8	29/29	19/19
1° tissues (1.5 months)								
Spleen					2/2	2/2	2/2	2/2
Thymus					2/2	2/2	2/2	2/2
Marrow					2/2	2/2	2/2	2/2
1° tissues (3 months)								
Spleen			1/1		2/2	2/2	2/2	2/2
Thymus					0/2	0/2	0/2	0/2
Marrow			1/1		2/2	2/2	3/3	2/3
2° CFU-S (1.5 months)	10/15	0/10*			3/28	0/3	13/43	0/13
2° CFU-S (3 months)			20/20	1/7	37/38	3/35	57/58	4/42
2° tissues (3 months)								
Spleen			2/2	1/2			2/2	1/2
Thymus			2/2	0/2			2/2	0/2
Marrow			2/2	1/2			2/2	1/2
3° CFU-S			7/29	0/7			7/29	0/7

*This set of 2° CFU-S was analyzed for RNA expression by reverse transcription/PCR.

and one for RNA analysis. Animals transplanted with 2–4 × 10^6 bone marrow cells were sacrificed after 1–3 months. Hematopoietic tissues were collected for nucleic acid analysis, and bone marrow cells were used directly to reconstitute a second generation of lethally irradiated female mice. Twelve days after the secondary BMT, the 2° CFU-S were isolated for DNA and RNA analysis. In one experiment, 3° BMT was performed from bone marrow of long-term-reconstituted 2° recipient animals in order to generate 3° CFU-S.

DNA and RNA Analysis. Genomic DNA was isolated by SDS/proteinase K and RNase digestion at 55°C for 3–4 hr. The digested tissues were extracted with phenol/chloroform; the DNA was precipitated in ethanol and resuspended in TE buffer. The presence of proviral GC sequences in the CFU-S and hematopoietic tissue samples was assayed by PCR using the human GC-specific oligonucleotide primers described by Weinthal *et al.* (5), followed by Southern blotting and hybridization with a ^{32}P-end-labeled internal oligonucleotide (8). Provirus DNA was also detected by Southern blot analysis after digestion of genomic DNA with the *Sst* II and *Xho* I restriction enzymes (BRL). These digestions release the 1.65-kb GC cDNA detected by hybridization with the 1.5-kb (*Sst* II/*Bam*HI) human GC cDNA probe. The probe was labeled with [^{32}P]dCTP by the random-priming method. Individual provirus integrants in the CFU-S and long-term hematopoietic tissues were detected by Southern blot analysis of genomic DNA digested with *Bam*HI, which cuts at one site in the provirus. Again, the Southern blot was hybridized with the 1.5-kb ^{32}P-labeled human GC cDNA probe.

RNA was isolated from the tissues by the acid guanidinium thiocyanate/phenol/chloroform method (11). RNA (15 μg) was electrophoresed on a 1.2% formaldehyde gel, denatured, neutralized, and transferred to a nylon membrane by capillary blotting. The filter was hybridized with the human GC cDNA probe. After a satisfactory exposure was obtained, the filter was stripped and rehybridized with the mouse β-actin DNA probe. For reverse transcription/PCR, 1 μg of RNA was reverse transcribed using the human GC-specific oligonucleotide primers, followed by PCR amplification of the cDNA as described above for the DNA samples.

Methylation Analysis. The methylation status of the proviral 5′ LTR in the CFU-S was determined by digestion of genomic DNA (15–25 μg) with *Bam*HI to reduce the size of the DNA fragments, followed by *Pvu* II digestion. The DNA was then precipitated with ethanol, redissolved in TE buffer, and divided into two equal portions, one of which was subjected to digestion with the methylation-sensitive enzyme *Sma* I. Completeness of the genomic DNA digestions was monitored by mixing a sample of the digestion mixture with either adenovirus type 2 DNA or λ DNA (BRL), which were subsequently run on a 1% gel. *Pvu* II and *Pvu* II/*Sma* I-digested DNA were electrophoresed and blotted to nylon membranes. The blots were probed with a ^{32}P-labeled fragment of the G2 vector from the *Spe* I site in the untranslated leader region to a *Pvu* II site near the 5′ end of the GC gene (see Fig. 3). Densitometric analyses were performed with the United States Biochemical SciScan 5000, measuring the relative densities of the 1.8-kb *Sma* I-resistant band and the 1.5-kb *Sma* I-sensitive band in each lane.

RESULTS

Expression of G2 *in Vivo* in Murine HSCs. Results of G2-mediated gene transfer and expression in the mouse

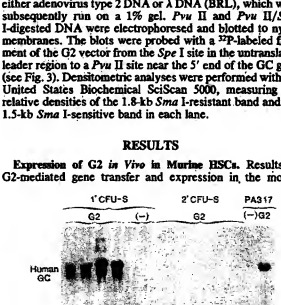


FIG. 1. Representative Northern blot analysis of the CFU-S generated in the mouse model of gene transfer/BMT (experiment 1 of Table 1). RNA from 1° CFU-S (lanes 1–6) and 2° CFU-S generated 1.5 months after primary BMT (lanes 7–11). Bone marrow cells used for BMT were transduced with G2 (lanes 1–4 and 7–11) or with the neomycin-containing control retroviral vector (lanes 5 and 6). RNAs from the fibroblast cell line PA317 (lane 12) and PA317 transduced with G2 (lane 13) were used as negative and positive controls, respectively, for GC mRNA. (*Upper*) Northern blot was probed with the human GC cDNA identifying the two proviral transcripts, the full-length 4.0 kb, and the spliced 3.5 kb. (*Lower*) Same blot was stripped and rehybridized with the mouse β-actin probe except for the control samples (lanes 12 and 13).

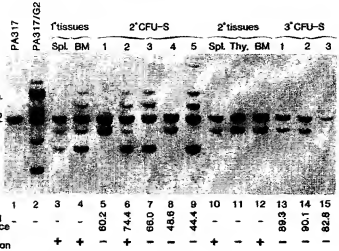


FIG. 2. Integration patterns of the G2 retroviral vector by Southern blot analysis. DNA was isolated from the tissues (experiment 2) and subjected to *BamHI* digestion, which cuts at one site in the retroviral vector. 1° tissues refer to the spleen (lane 3) and bone marrow (lane 4) of one animal sacrificed 3 months after primary BMT. The marrow of this animal was used to reconstitute 2° recipients to generate the 2° CFU-S depicted in lanes 5–9. One reconstituted 2° recipient was sacrificed 3 months after 2° BMT. The spleen, thymus, and bone marrow (2° tissues) were analyzed (lanes 10–12). The marrow from this 2° BMT recipient was used in a 3° transplant to generate the 3° CFU-S (lanes 13–15). The extent of methylation measured for each CFU-S is indicated as the percentage of *Sma* I resistance. Vector expression was measured by Northern blot analysis and is indicated as present (+) or absent (−).

model of gene transfer/BMT are shown in Table 1. In the 1° CFU-S, a gene transfer extent of 100% was recorded in the three experiments performed. Furthermore, RNA expression was detected at a high level in all nineteen 1° CFU-S examined. An example of RNA expression from the Mo-MuLV-proviral LTR is shown in Fig. 1. The two proviral RNA transcripts in the 1° CFU-S are depicted in lanes 1–4. These data confirm previous findings from our laboratory, which demonstrated high-level expression by G2 in 1° CFU-S (5).

To confirm hematopoietic reconstitution with genetically marked cells, hematopoietic tissues were analyzed from mice 1–3 months after primary BMT. The intact human GC cDNA was detected by genomic DNA Southern blot analysis in all 10 spleen and marrow samples and in the thymus from two of four primary recipient animals. Vector RNA transcripts were detected, by Northern blot analysis, in all tissues in which the provirus DNA sequences were present (Table 1).

Analysis of the 2° CFU-S showed a high overall retroviral gene transfer frequency, with 70 of 101 (70%) containing the G2 provirus. However, RNA expression was rarely detected in these tissues. In a total of three experiments, RNA expression from the MoMuLV-LTR was detected in only 4 of 55 (7.3%) of the 2° CFU-S that contained the G2 provirus. The Northern blot in Fig. 1 shows an example of five 2° CFU-S that did not express proviral RNA transcripts (lanes 7–11). Experiment 2 was carried out further by allowing two other secondary BMT recipients to survive for 3 months after the transplantation, producing tissues stably engrafted by the serially passaged marrow (Table 1). The spleen, thymus, and marrow of the two reconstituted 2° recipients contained G2 provirus DNA by Southern blot. Vector RNAs were detected at a low level in the spleen and marrow, but not the thymus, of one of the two animals. No vector transcripts were detected in the other reconstituted 2° recipient. Tertiary BMT was also performed in experiment 2 with the bone marrow obtained from the reconstituted 2° recipients 3 months after secondary BMT. Seven of the twenty-nine 3° CFU-S analyzed contained G2 provirus DNA; these were all derived from the donor 2° animal, which had shown expression of vector RNA in its tissues. None of the seven 3° CFU-S containing proviral DNA had detectable vector RNA expression.

Gene Transfer into Pluripotent HSCs. Provirus integration patterns were followed in 1°, 2°, and 3° recipient animals to demonstrate that the 2° and 3° CFU-S, in which the LTR was inactive, are descended from true pluripotent HSCs (Fig. 2). Both the spleen and marrow of the 1° recipient, 3 months after BMT (Table 1, experiment 2), showed the same pattern, with five proviral integrants (Fig. 2, lanes 3 and 4). The five 2° CFU-S that were produced from marrow of that primary recipient showed segregation of the vector integrants seen in the primary tissues into two different patterns. The first pattern, represented in lanes 6, 7, and 9, consisted of four proviruses (although there may be slight contamination of the sample in lane 6 with DNA from a CFU-S with the integrant seen in lane 5). Interestingly, despite the same pattern of integrants among these three 2° CFU-S, there was strong vector expression seen in the one in lane 6, whereas the two foci seen in lanes 7 and 9 had no detectable vector transcripts. The second pattern of vector integrants seen in two other 2° CFU-S (lanes 5 and 8) had one provirus. The latter single integrant of ≈4 kb was also detected in the spleen, thymus, and marrow of another reconstituted secondary animal 3 months after secondary BMT (lanes 10–12). Moreover, this integrant was transferred to three CFU-S of a 3° recipient animal. Thus, the same transduced stem cell that generated the 2° CFU-S was able to reconstitute all three hematopoietic tissues (spleen, thymus, and marrow) of a secondary animal sacrificed 3 months after BMT and also to generate 3° CFU-S in the 3° transplanted animal. The provirus present in the 2° and 3° CFU-S was transcriptionally inactive (Table 1).

Methylation Analysis of the CFU-S. Genomic DNA from transcriptionally active 1° CFU-S and from inactive 2° and 3° CFU-S were compared for *Sma* I resistance according to the protocol described in *Materials and Methods* and illustrated in Fig. 3. *Sma* I is a methylation-sensitive enzyme that can cleave DNA at the CCCGGG site only if the CpG sequence is not methylated; therefore, *Sma* I resistance is used as a measurement of DNA methylation. Two examples of the resulting Southern blots are shown in Fig. 4. In the packaging cell line PA317, the G2 provirus gives a 1.8-kb *Pvu* II band (Fig. 4, lanes 1), which is reduced to a 1.5-kb band upon *Sma* I digestion (lanes 2). Complete digestion by *Sma* I shows the absence of methylation of the 5′ LTR in the fibroblast cell

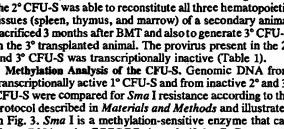


FIG. 3. Map of the G2 retroviral vector showing restriction sites and probe used for DNA methylation analysis.

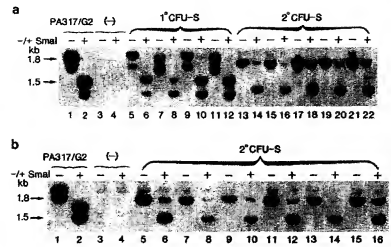


FIG. 4. Southern blot analysis of methylation of the G2 provirus in the 1° CFU-S (*a*) and 2° CFU-S (*a* and *b*). DNA was subjected to *Bam*HI and *Pvu* II with or without *Sma* I digestions. Lanes 3 and 4, labeled (−), represent DNA from normal control C57BL/6J mice. A second smaller band is seen in the PA317 cells and in the 1° CFU-S analyzed here due to the presence of a provirus with a short deletion in the 5′ untranslated region among the six copies in the packaging cell clone.

line. All *Sma* I digestions of DNA from 1° CFU-S show reduction in size of the vector-specific band to 1.5 kb, corresponding to lack of methylation of the *Sma* I site of the provirus 5′ LTR in 1° CFU-S. In contrast, the provirus in all 2° CFU-S shows some extent of *Sma* I resistance reflecting various degrees of methylation of the 5′ MoMuLV-LTR in these tissues.

To quantitate the degree of methylation at the *Sma* I site in the CFU-S, densitometric analyses were performed on the Southern blots, comparing the relative intensities of the 1.8- and 1.5-kb bands. Fig. 5 displays the percentage *Sma* I resistance of the G2 provirus in 3T3 fibroblasts, 1° CFU-S, and 2° and 3° CFU-S. In experiment 1 (Fig. 5*A*), the 1° CFU-S showed a mean of 8.9% *Sma* I resistance. In contrast, the average percentage *Sma* I resistance recorded in the 2° CFU-S generated 1.5 months after initial gene transfer/BMT was 24.6%, ranging from 9.0% to 47.6% over 10 samples analyzed. Thus, a 3-fold increase in percentage *Sma* I resistance was recorded between the 1° and 2° CFU-S of this experiment. The difference in methylation is statistically significant, with $P < 0.005$ by Student's *t* test analysis. Fig. 5*B* represents the comparison between 1° and 2° CFU-S derived 3 months after initial gene transfer/BMT in experiment 2. The percentage *Sma* I resistance was 14.9% in the 1° CFU-S and increased 3.8-fold to 56.8% in the 2° CFU-S. In the same experiment, the 3° CFU-S generated 3 months after secondary BMT showed 74.7% *Sma* I resistance, representing a 3.9-fold increase over the 19.3% *Sma* I resistance seen in other 1° CFU-S from the same experiment (Fig. 5*C*). Methylation of the provirus in both 2° and 3° CFU-S of this experiment was significantly greater than in 1° CFU-S ($P < 0.001$ by Student's *t* test). We conclude that the transcriptional inactivity observed in the 2° and 3° CFU-S is associated with methylation of the 5′ MoMuLV-provirus LTR at the *Sma* I site, 30 bp downstream of the transcription start site. Of note, the methylation status of the three 2° CFU-S with the common pattern of four vector integrants (Fig. 2, lanes 6, 7, and 9) was similar, ranging from 44% to 74%, despite the discordance for expression.

DISCUSSION

We have studied gene transfer of a MoMuLV-based retroviral vector into murine hematopoietic stem cells and expression from the MoMuLV-LTR promoter/enhancer elements in the progeny of the transduced cells. The expression from the proviral LTR was measured in the CFU-S derived from primary BMT performed after gene transfer, in the hematopoietic tissues of long-term reconstituted animals, and in the CFU-S generated after serial transplantation.

Our results demonstrate that the MoMuLV-LTR is a very efficient expression unit in 1° CFU-S. We have also detected expression from the MoMuLV-LTR in the hematopoietic tissues of transplant recipients 3 months after primary BMT.

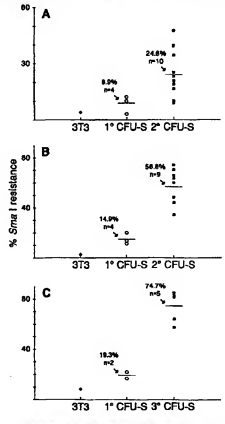


FIG. 5. Densitometric analysis to quantitate the extent of methylation of the MoMuLV-LTR in the 1°, 2°, and 3° CFU-S. The densities of the 1.8- and 1.5-kb bands after *Sma* I digestion were measured and percentage *Sma* I resistance was calculated. Each number represents the mean of three different readings. (*A*) Percentage *Sma* I resistance of samples from experiment 1. (*B* and *C*) Results from experiment 2 of the 2° CFU-S generated 3 months after secondary BMT and 3° CFU-S generated 3 months after 3° BMT, respectively. In each graph, the 1° CFU-S and either the 2° or the 3° CFU-S were analyzed from the same blot.

These results are comparable with previously published observations (4, 5, 12, 13). However, 1° CFU-S are mainly derived from committed progenitor cells restricted to the myeloid lineage (14, 15). Even the hematopoietic cells in the organs 3 months after BMT may be derived from progenitor cells capable of short-term but not long-term engraftment (16). By following the proviral integration pattern after serial transplantation, we were able to show that the cells capable of forming 2° CFU-S have the characteristics of long-lived, pluripotent HSCs. In this stringent assay of HSCs, we have observed that the MoMuLV-LTR is frequently inactive in the resultant progeny cells derived from the HSCs. We cannot, however, determine whether the absence of expression in the 2° CFU-S reflects the loss of expression by the LTR with time or whether the LTR is never active if inserted into the subclass of long-lived HSCs capable of producing 2° and 3° CFU-S.

The failure of transcription from the MoMuLV-LTR in hematopoietic tissues is in accord with prior observations (3–5, 17). However, the previous studies have mostly analyzed expression in reconstituted hematopoietic tissues of primary recipients. Moore *et al.* (17) have studied expression of human adenosine deaminase (ADA) by the MoMuLV-LTR vector expression through 2° CFU-S. Despite strong ADA expression in the primary recipients, they detected expression in only eighteen of seventy-two 2° CFU-S, although the percentage of these colonies that contained proviral DNA was not determined.

Methylation is associated with transcriptional inactivation of many genes and has specifically been seen in association with inactivity of the transduced MoMuLV-LTR in embryonic stem and embryonic carcinoma cell lines (7). Therefore, we examined the methylation status of the inactive 5′ MoMuLV-LTR in the 2° and 3° CFU-S. The presence of a high copy number of endogenous murine retroviral sequences creates a high background in the analysis of methylation across the provirus 5′ LTR. To overcome this problem, we used the restriction enzymes *Sma* I and *Pvu* II, which generate a specific provirus band detected by Southern blot analysis. Unfortunately, this assay restricted our analysis of methylation to one CpG dinucleotide contained in the *Sma* I site, 30 bp downstream from the transcription start site in the 5′ MoMuLV-proviral LTR.

Our analysis has indicated striking differences in the methylation patterns of this sequence. The *Sma* I site described is extensively methylated in the 2° CFU-S, which do not show MoMuLV proviral transcription, but is not methylated in the 1° CFU-S, which do express vector transcripts. We have observed that this same *Sma* I site is completely methylated in vector-transduced embryonic stem cell lines, which also do not show expression from the MoMuLV-LTR (data not shown). Moreover, studies done by Singer-Sam *et al.* (18) on the phosphoglycerate kinase promoter present on the inactive X chromosome have shown that methylation of a similar site, a *Hpa* II site (CCGG) at position +20, correlates with lack of transcription from the promoter. Nevertheless, the observed association between proviral methylation and expression inactivity does not show whether methylation plays a causal role in suppressing expression or is merely a secondary event after failure of expression has occurred.

Interestingly, the association between methylation and expression inactivity was not complete; at least one 2° CFU-S with extensive methylation at the *Sma* I site had a high level of vector transcripts. Other 2° CFU-S with the same proviral integrants failed to express the vector. This set of samples would indicate that neither integration site, methylation at the *Sma* I site, nor differentiation status of the transduced stem cell act to absolutely govern expression. The observed discordance suggests that commitment to expression or inactivity may stochastically be made after the HSC has differentiated to produce multiple pre-CFU-S.

The studies presented in this paper suggest that the wild-type MoMuLV-LTR may not be the ideal transcriptional unit for expression in pluripotent HSCs and their progeny cells. Successful approaches to overcoming this problem have been to use either modified LTRs with enhancer substitutions (19) or internal promoters derived from housekeeping genes (such as phosphoglycerate kinase or β-actin) instead of the wild-type MoMuLV-LTR (20, 21). Characterization of a transcriptionally active retroviral vector in HSCs may provide a better understanding of the regulation of gene expression occurring in these cells.

We would like to thank Dr. Robertson Parkman, Dr. Gay Crooks, and Dr. Punam Malik for carefully reading the manuscript; Dr. Wanda Krall for her helpful advice; and Dr. Ellen Bolotin for assistance in statistical analysis. The study was supported by National Institutes of Health Grant DK-42694 and March of Dimes Grant 1-FY92-0551. D.B.K. is a recipient of an American Cancer Society Junior Faculty Research Award (JFRA-320).

1. Mulligan, R. C. (1993) *Science* **260**, 926–932.
2. Palmer, T. D., Rosman, G. J., Osborne, W. R. & Miller, A. D. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1330–1334.
3. Williams, D. A., Orkin, S. H. & Mulligan, R. C. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2566–2570.
4. Kaleko, M., Garcia, J. V., Osborne, W. R. & Miller, A. D. (1990) *Blood* **75**, 1733–1741.
5. Weinthal, J., Nolta, J. A., Yu, X. J., Lilley, J., Uribe, L. & Kohn, D. B. (1991) *Bone Marrow Transplant.* **8**, 403–412.
6. Cedar, H. (1988) *Cell* **53**, 3–4.
7. Habers, K., Schnieke, A., Stuhlman, H., Jahne, D. & Jaenisch, R. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7609–7613.
8. Jaehner, D. & Jaenisch, R. (1985) *Nature (London)* **315**, 594–597.
9. Hoeben, R. C., Migchielsen, A. A. J., van der Jagt, R. C. M., van Ormondt, H. & van der Eb, A. J. (1991) *J. Virol.* **65**, 904–912.
10. Morgan, R. A., Cornetta, K. & Anderson, W. F. (1990) *Hum. Gene Ther.* **1**, 135–149.
11. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
12. Karlsson, S. (1991) *Blood* **78**, 2481–2492.
13. Krall, W. J., Challita, P. M., Perlmutter, L. S., Skelton, D. C. & Kohn, D. B. (1994) *Blood*, in press.
14. Ploemacher, R. E. & Brons, R. H. (1989) *Exp. Hematol.* **17**, 263–266.
15. Jones, R. J., Wagner, J. E., Celano, P., Zicha, M. S. & Sharkis, S. J. (1990) *Nature (London)* **347**, 188–189.
16. Lemischka, I. R., Raulet, D. H. & Mulligan, R. C. (1986) *Cell* **45**, 917–927.
17. Moore, K. A., Fletcher, F. A., Villalon, D. K., Utter, A. E. & Belmont, J. W. (1990) *Blood* **75**, 2085–2092.
18. Singer-Sam, J., Goldstein, L., Dai, A., Gartler, S. M. & Riggs, A. D. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1413–1417.
19. van Beusechem, V. W., Kukler, A., Einerhand, M. P. W., Bakx, T. A., van der Eb, A. J., van Bekkum, D. W. & Valerio, D. (1990) *J. Exp. Med.* **172**, 729–736.
20. Lim, B., Apperley, J. F., Orkin, S. H. & Williams, D. A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 8892–8896.
21. Guild, B. C., Finer, M. H., Housman, D. E. & Mulligan, R. C. (1988) *J. Virol.* **62**, 3795–3801.

X. RELATED PROCEEDINGS APPENDIX

No decisions have been rendered by a court or the Board in any proceeding identified pursuant to paragraph (c)(1)(ii) C.F.R. §41.37.

XI. CONCLUSION

For the foregoing reasons, Appellants respectfully submit that the Examiner's rejection of Claims 1-10, 12, 14, 18, 20, 21, 28, 30-34, and 41 is erroneous. Reversal of the rejections is respectfully requested. Appellants request that the Board render a decision as to the allowability of the Claims.

Respectfully submitted,

Dated: October 26, 2011 /John Mitchell Jones/

J. Mitchell Jones
Registration Number 44,174
Casimir Jones SC
2275 Deming Way
Suite 310
Middleton WI 53562